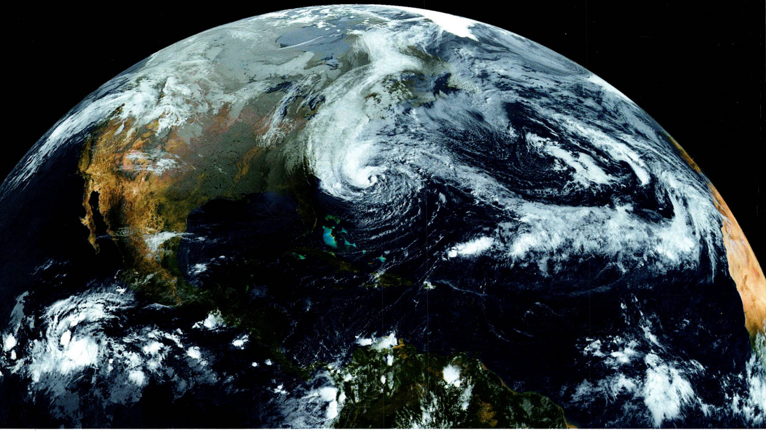
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

THE HEAT IS ON

A survival guide for the post-Kyoto world **PAGE 653**



CULTURE

DEAD RECKONING

Two exhibitions put death centre stage

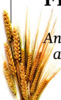
PAGE 671

GEOCHEMISTRY

MAKING THE SEABED

Global distribution of ocean-floor trace elements

PAGES 677 & 698



CEREALS

FEEDING THE BILLIONS

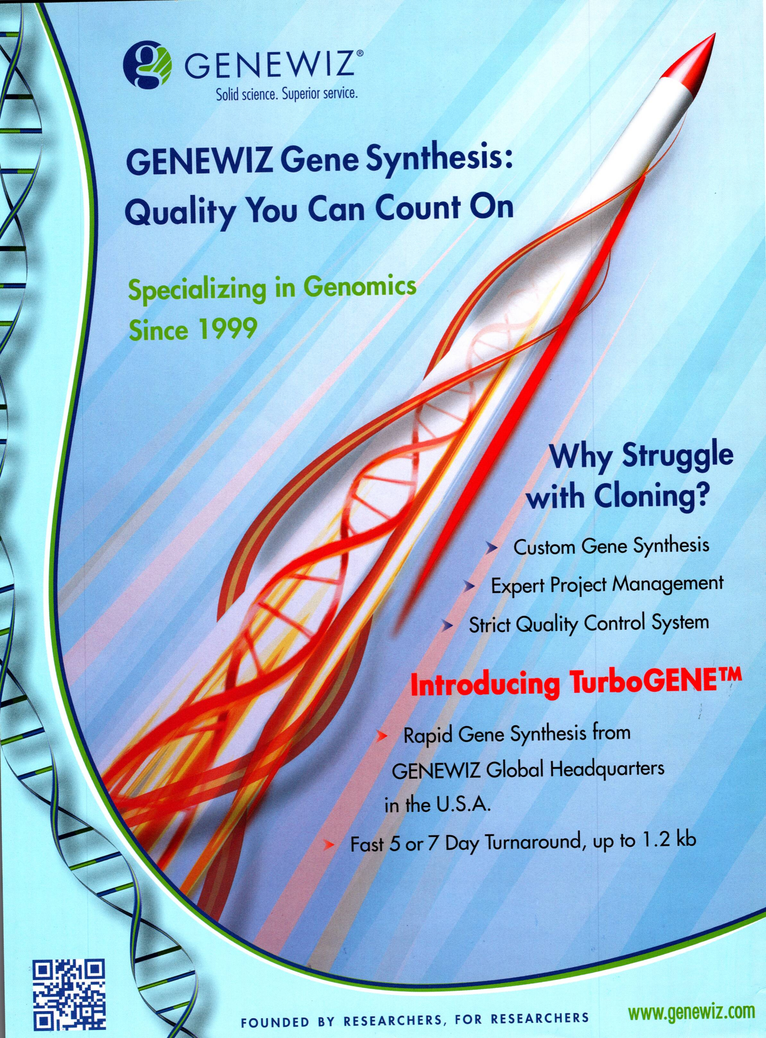
Analysis of bread wheat and barley genomes

PAGES 678, 705 & 711

➲ **NATURE.COM/NATURE**

29 November 2012 £10

Vol. 491, No. 7426



GENEWIZ®
Solid science. Superior service.

GENEWIZ Gene Synthesis:
Quality You Can Count On

Specializing in Genomics
Since 1999

Why Struggle with Cloning?

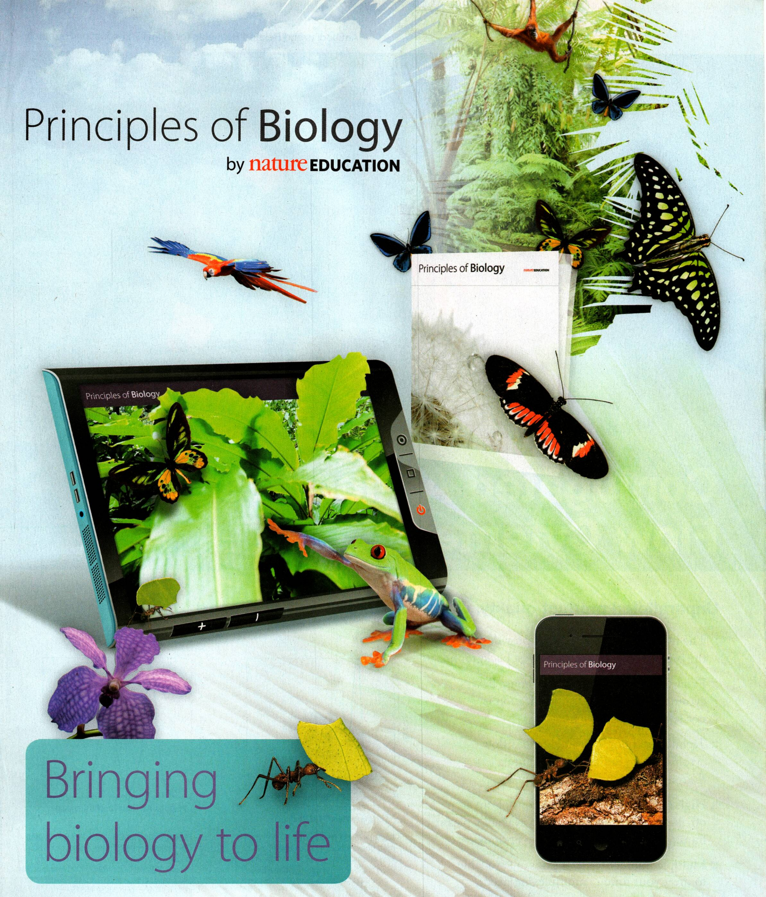
- Custom Gene Synthesis
- Expert Project Management
- Strict Quality Control System

Introducing TurboGENE™

- Rapid Gene Synthesis from GENEWIZ Global Headquarters in the U.S.A.
- Fast 5 or 7 Day Turnaround, up to 1.2 kb

FOUNDED BY RESEARCHERS, FOR RESEARCHERS

www.genewiz.com


Principles of Biology
Principles of Biology
Principles of Biology

Principles of **Biology**

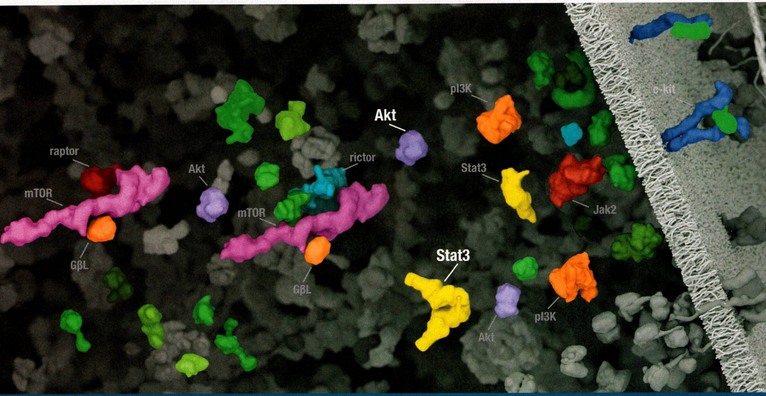
by nature **EDUCATION**

Bringing biology to life

To learn more about *Principles of Biology*, the first interactive textbook from Nature Education, visit **www.nature.com/principles**

nature publishing group npg

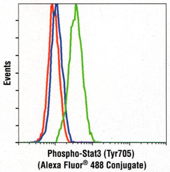
Antibodies and Related Reagents for Signal Transduction Research



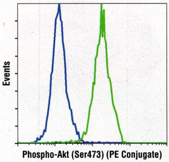
Conjugated Antibodies for Flow Cytometry *from Cell Signaling Technology*

Unparalleled product quality, validation, and technical support

- The highest quality antibodies provide you with the brightest signal and the lowest background.
- Extensive validation by our in-house flow cytometry group means optimization is not left up to you, the user.
- Technical support provided by our flow cytometry specialists translates into a thorough, fast, and accurate response.
- High quality custom conjugations of our off-the-shelf antibodies are also available, please inquire about sizing and pricing options.
- Multiple fluorochrome options are available to meet your unique multiplexing needs.



Phospho-Stat3 (Tyr705) (D3A7) XP® Rabbit mAb (Alexa Fluor® 488 Conjugate) #4323: Flow cytometric analysis of Jurkat cells, untreated (blue) or IFN-α treated (green), using #4323 compared to nonspecific control antibody (red).



Phospho-Akt (Ser473) (D9E) XP® Rabbit mAb (PE Conjugate) #5315: Flow cytometric analysis of Jurkat cells, untreated (green) or treated with LY294002 (PI3 Kinase Inhibitor) #9901, Wortmannin (PI3 Kinase Inhibitor) #9551, and U0126 (MEK1/2 Inhibitor) #9903 (blue), using #5315.

© 2012 Cell Signaling Technology, Inc. / XP® and Cell Signaling Technology® are trademarks of Cell Signaling Technology, Inc.
Alexa Fluor® is a trademark of Molecular Probes, Inc.

for quality products you can trust...
www.cellsignal.com

Cell Signaling
TECHNOLOGY®

Orders (toll-free) 1-877-616-2355 | **Technical support** (toll-free) 1-877-678-8324 support@cellsignal.com | **Inquiries** info@cellsignal.com | **Environmental Commitment** eco.cellsignal.com

For Research Use Only. Not For Use In Diagnostic Procedures.

CONTENTS

29 November 2012 / Vol 491 / Issue No 7426

AFTER KYOTO

FEATURES

653 **Introduction: After Kyoto**
As the Kyoto Protocol comes to a close, *Nature* looks to the future

654 **Awash in carbon**
More than ever, fossil fuels are the energy source of choice

656 **Hot air**
The legacy of a climate treaty

659 **No going back**
Many nations are laying plans to cope with the effects of a warming world

COMMENT

663 **The Kyoto approach has failed**
Dieter Helm
New technologies are key to combatting climate change

666 **Cap and trade finds new energy**
Michael Grubb
Emissions trading has not died with Kyoto

COVER: HURRICANE SANDY FROM SPACE (NASA GOES PROJECT)
COVER MINIPIC: ISTOCKPHOTO/THINKSTOCK

THIS WEEK

EDITORIALS

637 HEALTH
Misguided cancer goal
Making promises that can't be kept is not the best way to tackle disease

637 ENERGY
A way to buy time
Energy efficiency would help slow the effects of climate change

638 EUROPE
A bleak Horizon
Researchers must rally to fight cuts to Europe's research funding

WORLD VIEW

639 **Science should be ready to jump off 'the cliff'**
Colin Macilwain
Washington's much-hyped 'fiscal cliff' is just a bump in the road

RESEARCH HIGHLIGHTS

640 SELECTIONS FROM THE SCIENTIFIC LITERATURE
Marine-snail shells weakened / Blue whales on a roll / A yes to time reversal / Fixing damage due to cancer drugs

SEVEN DAYS

642 THE NEWS IN BRIEF
Polio initiative misses eradication goal / Coronavirus sparks concern / Europe in budget limbo / CO_2 hits record high / Grand Canyon gets flooded

NEWS IN FOCUS

645 FUNDING
Budget squeeze takes its toll on Europe's Earth-observation programme

646 ENERGY
Biofuel bottoms out in Brazil

647 EPIDEMIOLOGY
Keeping track of environmental exposures that cause disease

649 POLICY
Bulgarian research agency under fire

650 ATMOSPHERIC SCIENCE
Microsatellites seek climate data

CAREERS

785 COLUMN
When lab leaders take too much control
Charles Wood
Some of the latest management techniques are poison in the laboratory

786 CAREER BRIEFS

787 TURNING POINT
Glaciochemist Sarah Aciego takes a down-to-earth approach

COMMENT

BOOKS & ARTS

668 ENGINEERING
Turbulent genius
Allan McRobie

669 BOOKS IN BRIEF

670 HISTORY
Dreaming of the bomb
Istvan Hargittai

671 THANATOLOGY
Beyond the grave
Ewen Callaway

CORRESPONDENCE

672 Lost ash-tree lichens / Retiring chimps / Young scientists threatened by Europe's austerity drive

FUTURES

790
Transmission received
Peter J Enyeart

7,000 Human Proteins Made from Human Cells

More Assays!
Better Assays!

30% OFF*

The Largest Collection of Mammalian Cell-Derived Proteins

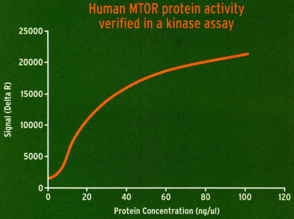
Protein functional studies

Compound screening

Immunogens for antibody production

Assay standards: ELISA, Mass Spec, and more

Bulk production available


Human MTOR protein activity verified in a kinase assay

*3000 proteins, expires 12/15/12
origene.com/protein

ORIGENE
Your Gene Company

CONTENTS

29 November 2012 / Vol 491 / Issue No 7426

RESEARCH

NEW ONLINE

673 Papers published this week at nature.com

NEWS & VIEWS

674 **COMPUTATIONAL MATERIALS SCIENCE**
Substitution with vision
Predicting alloy stability and elasticity
Gus L W Hart SEE LETTER P.740

675 **PLANT ECOLOGY**
Forests on the brink
Trees worldwide threatened by drought
Bettina M J Engelbrecht SEE LETTER P.752

677 **EARTH SCIENCE**
Magma chambers on a slow burner
Assessment of crystallization processes in ocean-floor magma chambers
Albrecht W Hofmann SEE ARTICLE P.698

678 **GENOMICS**
Decoding our daily bread
A detailed analysis of the hexaploid bread wheat genome
Peter Langridge SEE ARTICLE P.705

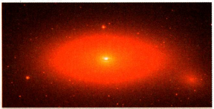
680 **IMMUNOLOGY**
Vitamins prime immunity
MR1 binds metabolites of B vitamins for presentation to MAIT cells
Wei-Jen Chua & Ted H Hansen
SEE ARTICLE P.717

681 **QUANTUM PHYSICS**
Strongly correlated transport
Tuning particle transport in a field-effect transistor
Lincoln D Carr & Mark T Lusk
SEE LETTER P.736

PERSPECTIVES

683 **CLIMATE SCIENCE Making sense of palaeoclimate sensitivity**
PALAEOSENS Project Members

692 **CLIMATE SCIENCE The mystery of recent stratospheric temperature trends**
D W J Thompson et al.



ASTRONOMY

Taking over

Over half the mass of this galaxy is found in the cental black hole. PAGE 729

ARTICLES

698 **EARTH SCIENCE The global pattern of trace-element distributions in ocean floor basalts**
H St C O'Neill & F E Jenner SEE N&V P.677

705 **GENOMICS Analysis of the bread wheat genome using whole-genome shotgun sequencing**
R Brenchley et al. SEE N&V P.678

711 **GENOMICS A physical, genetic and functional sequence assembly of the barley genome**
The International Barley Genome Sequencing Consortium

717 **IMMUNOLOGY MR1 presents microbial vitamin B metabolites to MAIT cells**
L Kjer-Nielsen et al. SEE N&V P.680

724 **BIOPHYSICS Visualizing transient low-populated structures of RNA**
E A Dethoff, K Petzold, J Chugh, A Casiano-Negroni & H M Al-Hashimi

LETTERS

729 **ASTRONOMY An over-massive black hole in the compact lenticular galaxy NGC 1277**
R C E van den Bosch et al.

732 **PLANETARY SCIENCES Active upper-atmosphere chemistry and dynamics from polar circulation reversal on Titan**
N A Teanby et al.

736 **PHYSICS Observing the drop of resistance in the flow of a superfluid Fermi gas**
D Stadler, S Krinner, J Meineke, J-P Brantut & T Esslinger SEE N&V P.681

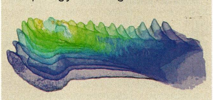
740 **MATERIALS SCIENCE A canonical stability–elasticity relationship verified for one million face-centred-cubic structures**
S B Maisel, M Höfler & S Müller
SEE N&V P. 674

744 **EARTH SCIENCE Rapid coupling between ice volume and polar temperature over the past 150,000 years**
K M Grant et al.

PALAEONTOLOGY

Oral history

Early teeth and early jaws have surprisingly varied origins. PAGE 748



748 **PALAEONTOLOGY Development of teeth and jaws in the earliest jawed vertebrates**
M Rücklin et al.

752 **PLANT ECOLOGY Global convergence in the vulnerability of forests to drought**
B Choat et al. SEE N&V P.675

756 **EVOLUTION The genomic landscape of species divergence in *Ficedula* flycatchers**
H Ellegren et al.

761 **NEUROSCIENCE A map of visual space in the primate entorhinal cortex**
N J Killian, M J Jutras & E A Buffalo

765 **STEM CELLS Self-renewal of embryonic-stem-cell-derived progenitors by organ-matched mesenchyme**
J B Sneddon, M Borowiak & D A Melton

769 **IMMUNOLOGY Impaired intrinsic immunity to HSV-1 in human iPSC-derived TLR3-deficient CNS cells**
F G Lafaille et al.

774 **VIROLOGY Resurrection of endogenous retroviruses in antibody-deficient mice**
G R Young et al.

779 **STRUCTURAL BIOLOGY Structure of the chemokine receptor CXCR1 in phospholipid bilayers**
S H Park et al.

784 **CORRIGENDUM Yeretssian *et al.* reply**
G Yeretssian et al.

784 **CORRIGENDUM CD95 promotes tumour growth**
L Chen et al.

nature

29 November 2012 / Vol 491 / Issue No. 7426

EDITORIAL

LONDON, MUNICH, WASHINGTON DC, NEW YORK, BOSTON, PARIS, SAN FRANCISCO, TOKYO **nature@nature.com**

EDITOR-IN-CHIEF: Philip Campbell

EXECUTIVE EDITOR: Nick Campbell

PUBLISHING EXECUTIVE EDITOR: Maxine Clarke **EXECUTIVE EDITOR:** Véronique Kiermer

CHIEF MAGAZINE EDITOR: Tim Appenzeller

THIS WEEK/NEWS/FEATURES: Alison Abbott, David Adam, Ananyo Bhattacharya, Geoff Brumfiel, Declan Butler, Ewen Callaway, Erika Check Hayden, Daniel Cressey, David Cyranoski, Natasha Gilbert, Eric Hand, Heidi Ledford, Corie Lok, Brendan Maher, Emma Marris, Richard Monastersky, Helen Pearson, Mark Peplow, Eugenie Samuel Reich, Ivan Semeniuk, Quirin Schiermeier, Jeff Tollefson, Richard Van Noorden, Meredith Wadman, M Mitchell Waldrop **COMMENT:** Sara Abdulla, Joanne Baker, Rosalind Cotter, Barbara Kiser, Alison McCook, Lucy Odling-Smee **NATURE PODCAST/VIDEO:** Adam Rutherford, Thea Cunningham, Kerri Smith, Charlotte Stoddart **NEWS AND VIEWS:** Sadaf Shadan, Ana Lopes, Andrew Mitchinson, Marian Turner **RESEARCH:** Ritu Dhand, Karl Ziemelis, Francesca Cesari, I-han Chou, Tanguy Chouard, Rosamund Daw, Alex Eccleston, Angela Eggleston, Joshua Finkelstein, Henry Gee, Patrick Goymer, Noah Gray, Marie-Thérèse Heemels, Magdalena Helmer, Karen Howell, Barbara Marte, Andrew Mitchinson, Juliane Mössinger, Deepa Nath, Leslie Sage, Magdalena Skipper, Clare Thomas, Maria Trajkovska, John VanDecar, Liesbeth Venema, Ursula Weiss, Michael White **INSIGHTS/REVIEWS/PERSPECTIVES:** Ursula Weiss **CAREERS:** Gene Russo, Karen Kaplan **TECHNOLOGY FEATURES:** Vivien Marx **FUTURES:** Colin Sullivan **SUPPLEMENTS:** Herb Brody, Michelle Grayson, Tony Scully **SUBEDITORS:** Colin Sullivan, Mary Abraham, Sarah Archibald, Nicola Bailey, Nicolas Fanget, Isobel Flanagan, Paul Fletcher, Anne Haggart, Sasha Lewis, Dinah Loon, Anna Novitzky, Yukie Ozawa, Katherine Perry, David Price, Jenny Rooke, David Winters, Anna York **PRODUCTION:** Jenny Henderson, Leonora Dawson-Bowling, Dan Foley, Stewart Fraser, Susan Gray, Simon Gribbin, Hazel Mayhew, James McQuat, Emilia Orviss, Yvonne Strong, Robert Sullivan, Roberta Tarran, Charles Wenz **ART AND DESIGN:** Kelly Buckheit Krause, Wesley Fernandes, Madeline Hutchinson, Barbara Izdebska, Paul Jackman, Jasiek Krzysztofiak, Annthea Lewis, Chris Ryan, Nik Spencer, Claire Welsh, Mark Wood **JAPAN EDITION:** Eiji Matsuda, Mika Ishida, Mitsuyo Utsugi, Sayaka Ayame **PRESS OFFICE:** Rachel Twinn, Neda Afsarmanesh, Lisa Boucher, Jenny Chapman, Rebecca Walton **ADMINISTRATION:** Paul Sng, Liz Bankes, Roseann Campbell, Peter Cary, Ffion Cleverley, Dana Coles, Faye Fornasier, Ashlee Gagui, Peter Hewitt, Caroline McLean, Anastasia Panoutsou, Kenneth Simpson, Kimberly Tolleson, Michelle Viscuso, Diane Yorke

PUBLISHING

LONDON **feedback@nature.com**

MANAGING DIRECTOR & PUBLISHER: Steven Inchcoombe

ASSOCIATE PUBLISHER: Arash Hejazi

PUBLISHING PROJECT MANAGERS: Claudia Deasy, Christian Manco

TOKYO **feedback@natureasia.com**

REGIONAL MANAGING DIRECTOR — ASIA-PACIFIC: David Swinbanks

DIRECTOR — ASIA-PACIFIC: Antoine E Bocquet

OPERATIONS DIRECTOR — ASIA-PACIFIC: Hiroshi Minemura

DISPLAY ADVERTISING, SPONSORSHIP & NATUREJOBS

GLOBAL HEAD: Andrew Douglas a.douglas@nature.com

DISPLAY ADVERTISING AND SPONSORSHIP

GLOBAL HEAD: Gerard Preston g.preston@nature.com

HEAD OF SPONSORSHIP: Reya Silao Tel: +44 (0)20 7843 4977 (r.silao@nature.com)

EUROPE/REST OF WORLD display@nature.com Tel: +44 (0)20 7843 4960

GERMANY/SWITZERLAND/AUSTRIA: Sabine Hugi-Fürst Tel: +41 52761 3386

UK/IRELAND/SCANDINAVIA/SPAIN/PORTUGAL: Evelina Rubio Hakansson Tel: +44 (0)20 7014 4079

FRANCE/BELGIUM/THE NETHERLANDS/LUXEMBOURG/ITALY/OTHER EUROPE: Nicola Wright Tel: +44 (0)20 7843 4959

NORTH AMERICA **display@natureny.com**

NEW ENGLAND/EASTERN CANADA: Sheila Reardon Tel: +1 617 475 9202

NEW YORK/MID-ATLANTIC/SOUTHEAST: Jim Breault Tel: +1 212 726 9334

MIDWEST: Mike Rossi Tel: +1 212 726 9255

WEST COAST/WESTERN CANADA: George Lui Tel: +1 415 781 3804

NORTHWEST: Kayla McCutchan Tel: +1 415 403 9038; Fax: +1 415 781 3805

ASIA-PACIFIC **display@natureasia.com**

ASIA-PACIFIC HEAD: Kate Yoneyama

JAPAN: Ken Mikami Tel: +81 3 3267 8765

GREATER CHINA/SINGAPORE: Gloria To Tel: +852 2811 7191

INDIA: Vikas Chawla Tel: +91 124 288 1057

SPONSORSHIP

sponsorship@nature.com

SENIOR BUSINESS DEVELOPMENT MANAGER: David Bagshaw Tel: +1 212 726 9215

BUSINESS DEVELOPMENT MANAGER: Greg Valero Tel: +1 212 726 9360

BUSINESS DEVELOPMENT MANAGER: Will Piper Tel: +44 (0)20 7014 4181

NATUREJOBS

naturejobs@nature.com

EUROPEAN HEAD OFFICE, LONDON: Tel: +44 (0)20 7843 4961

GLOBAL HEAD OF CLASSIFIED ADVERTISING: Nils Moeller (n.moeller@nature.com) **BUSINESS DEVELOPMENT MANAGER, MIDDLE EAST:** Jon Giuliani (j.giuliani@nature.com) **ADVERTISING PRODUCTION MANAGER:** Stephen Russell **NATUREJOBS ONLINE PRODUCTION:** Dennis Chu **US HEAD OFFICE, NEW YORK:** naturejobs@us.nature.com **US SALES MANAGER:** Ken Finnegan **JAPAN HEAD OFFICE, TOKYO:** Tel: +81 3 3267 8765 **ASIA-PACIFIC HEAD:** Kate Yoneyama **ASIA-PACIFIC SALES & BUSINESS DEVELOPMENT MANAGER:** Yuki Fujiwara **BUSINESS DEVELOPMENT MANAGER, GREATER CHINA, SINGAPORE:** Gloria To Tel: +852 2811 7191 **INDIA:** Vikas Chawla Tel: +91 1242881057 (v.chawla@nature.com)

MARKETING & SUBSCRIPTIONS

USA/CANADA/LATIN AMERICA **subscriptions@us.nature.com**

Tel: (USA/Canada)+1 866 363 7860; (outside USA/Canada) +1 212 726 9223

MARKETING: Sara Girard **FULFILMENT:** Karen Dawson

JAPAN/CHINA/KOREA **subscriptions@natureasia.com**

Tel: +81 3 3267 8751 **MARKETING:** Sachiyo Ikeda **FULFILMENT:** Masayo Tanaka

EUROPE/REST OF WORLD **subscriptions@nature.com**

Tel: +44 (0)1256 329242

MARKETING: Elena Woodstock, Marcus Parker, Hannah Phipps **FULFILMENT:** Jean Burrows

INDIA **npgindia@nature.com**

Tel: +91 124 2881053/54

HEAD OF BUSINESS DEVELOPMENT, INDIA: Debashish Brahmachari **MARKETING:** Harpal Singh Gill

ANNUAL SUBSCRIPTIONS (UK/Europe, including post and packing)

FULL RATE: £1,400 **PERSONAL RATES:** UK £135; Europe €209; Rest of World £210

STUDENT RATES: UK £79; Europe €122; Rest of World £125

Individual rates apply to payments by personal cheque or credit card. Back issues: £20.00.

ANNUAL SUBSCRIPTIONS (USA, including post and packing)

INSTITUTIONAL/CORPORATE RATE: $2,730 ; **PERSONAL RATE:** $199; **STUDENT RATE:** $99; **POSTDOC RATE:** $119

Printed in USA. Individual rates available only to subscribers paying by personal check or credit card. Orders for student/postdoc subscriptions must be accompanied by a copy of student ID. Rates apply to USA, Canada, Mexico/Central & South America. Add 7% GST tax in Canada (Canadian GST number 140911595). **BACK ISSUES:** US$20.00.

ANNUAL SUBSCRIPTIONS (Asia, including post and packing)

JAPAN: **年間購読料金**: *Nature* のみ 536,300 円 +26,815(消費税)=563,115円(合計)
Nature + *Nature Digest* 574,500 円 +28,725(消費税)=603,225円(合計)
お問い合わせ：日本出版貿易株式会社 商品統括部 雑誌営業課
〒101-0064 東京都千代田区猿楽町1-2-1 Tel: 03-3292-3767 Fax: 03-3292-8766
個人/学生購読について： ご本人名義で個人的に支払い、私物として購読する場合に適用されます。
NPG ネイチャー アジア・パシフィック　フルフィルメント
〒162-0843 東京都新宿区市谷田町2-37 千代田ビル Tel: 03-3267-8751Fax: 03-3267-8752

CHINA: www.natureasia.com/ch/subscribe; **KOREA:** www.natureasia.com/korea/subscribe

NATURE ONLINE

WWW.NATURE.COM/NATURE

CHIEF TECHNOLOGY OFFICER: Howard Ratner **ASSOCIATE DIRECTOR, NATURE.COM:** Daniel Pollock **PLATFORM SERVICES:** Jeremy Macdonald **WEB PRODUCTION:** Alexander Thurrell, Dipti Shah **WEB DESIGN:** Tom Scott, Juan-Carlos Sobrino, Luke Stavenhagen, Cianna Timbers **PLATFORM INTERFACES:** Amanda Ward

SITE LICENCES, FULFILMENT & CUSTOMER SERVICES

SITE LICENCES: npg.nature.com/libraries feedback@nature.com

CHIEF OPERATIONS OFFICER: John Carroll **ASSOCIATE DIRECTOR OF OPERATIONS:** Dominic Pettit

CUSTOMER SERVICE: Liz Davidson **ORDER PROCESSING OPERATIONS:** Steve Poulter

NATURE PUBLISHING GROUP OFFICES

LONDON: The Macmillan Building, 4 Crinan Street, London N1 9XW Tel: +44 (0)20 7833 4000 Fax: +44 (0)20 7843 4596/7

BASINGSTOKE: Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants RG21 6XS, UK Tel: +44 (0)1256 329242

WASHINGTON DC: 968 National Press Building, 529 14th St NW, Washington DC 20045-1938 Tel: +1 202 737 2355

NEW YORK: 75 Varick St, 9th Floor, New York, NY 10013-1917 Tel: +1 212 726 9200

BOSTON: 25 First Street, Suite 104, Cambridge, MA 02141 Tel: +1 617 475 9275

SAN FRANCISCO: 225 Bush Street, Suite 1453, San Francisco, CA 94104 Tel: +1 415 403 9027

MUNICH: Nymphenburger Strasse 14, D-80335 München Tel: +49 89 549057-13

PARIS: Moulin de Malicorne, Sainte Maure de Touraine, Paris 37800 Tel: +33 2 47 35 72 15

TOKYO: Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843 Tel: +81 3 3267 8751

INDIA: Nature Publishing Group, 3A, 4th Floor, DLF Corporate Park, Gurgaon 122002 Tel: +91 124 2881053/54

Nature® (ISSN 0028-0836) is published by Nature Publishing Group, a division of Macmillan Publishers Ltd (The Macmillan Building, 4 Crinan Street, London N19XW). Registered as a newspaper at the British Post Office. North and South American orders to subscriptions@natureny.com (Tel: +1 866 363 7860). Other orders to subscriptions@nature.com (Tel. +44 (0)1256 329242; Fax +44 (0)1256 812358). Authorization to photocopy material for internal or personal use, or internal or personal use of specific clients, is granted by *Nature* to libraries and others registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided the relevant copyright fee is paid direct to CCC, 222 Rosewood Drive, Danvers MA 01923, USA. Identification code for *Nature*: 0028-0836/03. CPC PUB AGREEMENT #40032744. In the US, *Nature* (ISSN 0028-0836) is published weekly on Thursday, except the last week in December by Nature Publishing Group, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. US Periodicals postage paid at New York NY, and additional mailing post offices. US POSTMASTER: send address changes to *Nature*, Subscription Dept, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. Published in Japan by NPG Nature Asia-Pacific, Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843, Japan. © 2012 Macmillan Publishers Limited. All rights reserved.

THIS WEEK

EDITORIALS

IMPASSE IN BRUSSELS Why the budgetary deadlock is bad news for research **p.638**

WORLD VIEW The US fiscal cliff isn't as fearsome as it looks **p.639**

SLOW TAKE-OFF Weak feathers hampered early birds **p.641**

Misguided cancer goal

An influential US advocacy group has set a deadline to beat breast cancer by 2020. But it puts public trust at risk by promising an objective that science cannot yet deliver.

Hope is not a good strategy, in life or in disease research. So the setting of goals, and the drive to reach them, is to be commended, and cancer is no exception. But a 2020 deadline for 'ending' breast cancer that former US President Bill Clinton endorsed earlier this month is misguided. Like other 'beat cancer' deadlines that are regularly floated, it is potentially harmful to the public trust that underpins the whole research enterprise, not to mention to the patients who understandably cling to hope, whatever its validity.

Clinton, who lost his mother to breast cancer, has become honorary chairman of a two-year-old campaign by the National Breast Cancer Coalition, which declares on its website that it has "One Mission: To End Breast Cancer by January 1, 2020". The advocacy and research-funding organization, based in Washington DC, adds that it has a "strategic plan" to achieve that mission, by focusing on prevention and on eliminating the metastatic form of the disease, which is what kills.

The coalition provides a 4.5-page "blueprint" that is long on aspiration and short on scientific detail. For instance, it declares that by 2020 "we must understand how to prevent people from getting breast cancer in the first place". This goal leans heavily on the development of a preventive breast-cancer vaccine. A research plan for this is said to be "in place" and will serve as a model for other, "catalytic projects". These could include exploiting the role of viruses and inflammation in breast cancer, and targeting the immune system to prevent metastasis.

Ambitious goals are perfectly defensible, and indeed desirable, when we have the means to achieve them. The campaign to eradicate smallpox made eminent sense once a vaccine was ready, as does the goal of eliminating polio. Yet the thorny problems of finishing off even polio, for which we have had a vaccine for nearly 60 years, provide a cautionary tale about the advisability of setting out to eliminate any disease.

This is particularly true of the myriad diseases we collectively call cancer, the complexities of which we have scarcely begun to fathom. Consider just one study, published earlier this year (P. J. Stephens *et al. Nature* **486,** 400–404; 2012), which analysed protein-coding genes in breast cancers from 100 different women and found no fewer than 40 different mutational drivers of the disease. These were found in 73 different combinations in the 100 patients, who each had between one and six mutations. The low-hanging fruit here is scarce: only 28 of the patients harboured just one mutation, and finding a targeted therapy for even these single-mutation cases will be a daunting task.

Added to that is the disease's intractability. It cannot be banished like smallpox; our biologies are by definition vulnerable to a disease that has infinite manifestations profoundly rooted in our genetics. Even if a panoply of promising therapies were available, the eight to ten years it takes to complete a clinical trial makes a 2020 deadline impossible. As for prevention, truly valuable trials require not years but decades, because of the various influences on breast-cancer development during a lifetime. Britain's Breakthrough Generations Study, which recruited its 100,000th participant in 2009, anticipates running for 40 years.

The National Breast Cancer Coalition counters that such arguments cater to those content with the status quo — what the coalition sees as the drift of a research enterprise that, after decades of investment, is not motivated by sufficient urgency. On the contrary: we are all for urgency, but in the service of goals that are within the realms of possibility.

Here are a few. Set out to identify all tumours in which the *HER2* gene is mutated and treat them with the drug Herceptin (trastuzumab) by 2020. The treatment is known to work for this genetic category of the disease, so this is not inconceivable. Or declare that in five years, we will have developed several robust breast-cancer models that could rapidly be deployed to evaluate the functional significance of the mutations and polymorphisms that genomics is uncovering at a breathtaking rate. A project such as this, with finite parameters and price tag, can be pegged to an achievable time frame.

> ***"Discovery does not answer to deadlines."***

Or, tackle another cancer afflicting women by campaigning to overcome the apathy with which the human papillomavirus vaccine has been greeted in the United States. Universal vaccination of 11- and 12-year-old girls against the cervical-cancer-causing virus would, at a stroke, provide huge gains against the roughly 4,000 deaths and 12,000 new cases of this cancer that are seen in the United States each year.

Discovery does not answer to deadlines, and campaigns that pretend that it does risk wasting public trust, whether from the taxpayers who support the US National Institutes of Health or from the millions of donors who give to dozens of disease-advocacy groups. There is a fine line between creating a sense of urgency and promising too much; it is best to stay on the side of the line that is realistic about how science works, and about what is currently achievable. ■

A way to buy time

With climate talks inching along, gains in energy efficiency could slow the rise in emissions.

This week and next, diplomats from around the world gather once again to discuss global warming. With commitments under the Kyoto Protocol ending this year (see page 653), one key goal of the United Nations meeting in Doha is to make progress towards the 2015 signing of a new global climate treaty, to take effect by 2020. The world is on track for a temperature increase of up to 4 °C by the end of the century, but the UN hopes to limit that to just 2 °C.

Unfortunately, diplomacy and global warming operate on incompatible schedules. An eight-year wait for action would seem to put the warming goal firmly out of reach. But there are ways to buy time for global diplomacy, and energy efficiency is at the top of the list.

The *World Energy Outlook 2012* report from the International Energy Agency (IEA) suggests that the global infrastructure could lock in enough carbon emissions by 2017 to exceed the 2 °C goal, unless facilities such as power plants, factories and buildings are expensively retrofitted or prematurely retired. But the IEA found that improving energy efficiency could give the world another five years to change course and begin the transition to renewables and other low-carbon energies.

Globally, energy use is projected to increase by more than one-third by 2035, despite promises by Japan, Europe, China and the United States to curb demand. In an 'efficient world' scenario, with more countries embracing bigger efficiency goals, the projected energy demand could be cut by half. For perspective, the IEA estimates that the modest efficiency increases achieved between 1980 and 2010 reduced global energy demand by 35% — roughly equivalent to the energy currently consumed by China and the United States combined.

The IEA suggests that more-aggressive efficiency measures, such as a broad shift toward efficient appliances, vehicles, homes and factories, would cost an extra US$11.8 trillion between now and 2035. But the pay-off would be substantial: direct fuel expenditures would fall by $17.5 trillion, and investments in energy infrastructure by nearly $5.9 trillion. Those savings would be reinvested elsewhere, helping to increase global economic output by some $18 trillion. Unfortunately, the potential gains are dispersed throughout a complex marketplace that tends to reward short-term thinking.

Governments must pursue solutions at all levels, and not wait until the next global treaty. Reducing subsidies on fossil fuels would cut energy consumption, for instance, as would increasing consumption taxes. High energy taxes help to explain why Japan and Europe are leaders in energy efficiency, just as increasing oil prices on the global market have encouraged Americans to reduce their oil consumption.

But playing with the price won't work if the signals aren't reaching the right people. Buildings are responsible for roughly one-third of global greenhouse-gas emissions, but builders have no incentive to invest in energy-efficient technologies if tenants and owners will foot the energy bill. To change that, governments can strengthen building codes for new construction and create financial incentives that reduce the up-front costs of retrofitting. They can also require energy audits when properties are sold; this encourages buyers and sellers alike to consider long-term operating costs.

"Improving energy efficiency could give the world another five years to change course."

In Doha and beyond, negotiators must look for opportunities for the world to embrace new and more ambitious climate goals. At the same time, governments must do everything they can to follow through with their own climate commitments, reduce carbon footprints at home and lay the groundwork for future steps. Stabilizing the climate will require monumental efforts on all fronts, and governments should recognize that money spent now on curbing greenhouse-gas emissions is a long-term investment that will pay off down the road. Nowhere is this clearer than in the arena of energy efficiency. ■

A bleak Horizon

Researchers should lobby against heavy cuts to pan-European research funds.

After much posturing and politicking, European leaders walked away from talks last week without a deal on the European budget for the rest of the decade. The breakdown casts into limbo a European Commission proposal to apportion around €80 billion (US$104 billion) to research over the period 2014–20 — a €29.5 billion rise on Europe's current seventh Framework programme. And it augurs trouble for research when the impasse is finally broken.

With 27 nations each pushing for their own priorities, finding an agreement on spending plans is inevitably complex, and the tight economic climate aggravated the differences even more than usual.

The key divisive factor is the demand from wealthy nations, including the United Kingdom, Germany and the Netherlands, for substantial cuts to the total €1.025-trillion European Union (EU) budget — a rise of around €50 billion on spending between 2007 and 2013 — proposed by the European Commission. Early in the talks, European Council president Herman Van Rompuy, who is chairing the negotiations, proposed a cut of €80 billion. Media reports say that rich nations are looking for further cuts, of between €30 billion and €75 billion. Speaking to reporters after the talks broke down on Friday afternoon, Van Rompuy said that member states had found a "sufficient degree of potential convergence" to make an agreement on the budget possible early next year.

This should leave enough time for the European Parliament, member states and the commission to thrash out the final details of the research programme, known as Horizon 2020, just in time for research projects to start in 2014, as planned. But that is one of the few bright spots in the outlook for research.

Of the cuts suggested by Van Rompuy, the Horizon 2020 research programme comes out among the worst, with a proposed 12% reduction in funding, according to calculations by the Initiative for Science in Europe (ISE), an independent advocacy coalition of learned societies and scientific organizations in Heidelberg, Germany. The Galileo satellite network, set to rival the US Global Positioning system, faces a 10% cut, and the budget for ITER, the world's largest nuclear-fusion experiment, is also under threat. Van Rompuy says member states agree that the final budget should encourage economic growth, by focusing spending on research and innovation, as well as on jobs. But EU politics force other priorities. The sharp cuts for research in the Van Rompuy plans allow for more moderate reductions of 3.7% in the budget for agriculture to appease France, and of 5.6% to 'cohesion funds' meant for poorer EU regions, to bring Poland on board with the negotiations.

If the proposed 12% cut to research funding sticks in the final deal, all aspects of the Horizon 2020 programme are likely to suffer equally. Unforgivably, this would include the programme's 'Excellent Science' initiatives, such as the European Research Council (ERC), which funds investigator-led frontier research, as well as research infrastructures, such as CERN — the world's largest particle-physics laboratory, near Geneva in Switzerland, and the institution responsible for the recent discovery of the Higgs boson. The valuable Marie Curie fellowships through which young researchers gain support for career development and experience working in labs abroad would also be threatened.

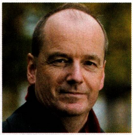
Helga Nowotny, president of the ERC, sees a bleak future for the council under the Van Rompuy proposals. She fears that the suggested cuts could result in funding for grants in 2014 dropping below levels available in 2009–10. Reductions of this magnitude will decimate success rates, particularly for young researchers, for whom other funding sources are scarce, she says. This would seriously damage the reputation painstakingly built by the ERC since it was founded just five years ago.

European researchers should do everything in their power to articulate the case for Europe's developing excellence, on which its future supply of scientific and technical manpower will depend. They should lobby their national leaders and support the efforts of the ISE. They can start by signing the petition, which had, as *Nature* went to press, collected almost 149,000 signatures, at: go.nature.com/s2nm1w. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

WORLD VIEW

A personal take on events



Science should be ready to jump off ‘the cliff’

Researchers can find plenty to like in a US budget scenario that scientific societies are comparing to the apocalypse, says **Colin Macilwain**.

There’s another warning note in my in-box this week. It is the latest in a long line of messages from US biologists’ main lobby group, the US Federation of American Societies for Experimental Biology (FASEB), about the “devastating” implications if the country careers off the ‘fiscal cliff’ in January.

The fiscal cliff is a set of sharp budget cuts (called sequestration) and tax increases that will take effect in January if Congress and the White House fail to agree before then on other ways to balance the budget.

Now, what I want to know is why science lobbyists in Washington DC have spent all summer panicking publicly about a budget plan that many of the people they represent would consider the least-bad outcome — for both US society and US science — of those on the menu.

The United States, most observers agree, faces an outlandish deficit. This year, US$3.6 trillion (24% of gross domestic product) will be spent by the government, but only $2.2 billion will be raised in taxes. Scientists know as well as anyone that this is unsustainable.

Wrangling over how to tame the deficit ended a year ago, when a congressional ‘super-committee’ failed to reach agreement. That left what is now known as the fiscal cliff — a fall-back arrangement agreed in August 2011 to force a better deal. It mandates that unless alternative plans are agreed, taxes will rise and across-the-board spending cuts will take effect.

For those of a progressive bent, the fiscal cliff has many attractions. First, it spreads cuts evenly across all ‘discretionary’ spending — including the half that goes to the Pentagon. Second, it protects Social Security, Medicare and Medicaid — the linchpins of the United States’ threadbare welfare state — from any cuts whatsoever. In so doing, it refuses to balance the budget on the backs of the poor.

FOR THOSE OF A PROGRESSIVE BENT, THE FISCAL CLIFF HAS MANY ATTRACTIONS.

Finally, and most importantly, it closes 80% of the deficit through higher taxation, and only 20% through spending cuts. That’s a sensible approach in a country where income tax rates — on the middle class as well as on the rich — have grown unfeasibly low.

The fiscal cliff, then, is a tough budget package that leans firmly to the left. How did a right-leaning Congress get there? Well, lawmakers never thought it would get enacted. Now they are trying to unravel it. And every special interest in Washington DC, from FASEB to the National Association of Manufacturers, is keen to lend a hand.

The scientific societies have been warning all summer that sequestration would be a disaster for science, imposing cuts of up to 8% in the budgets for 2013. Under this scenario, the National Institutes of Health would, if past is prelude, reduce its average annual grant from about $450,000 to $400,000 — not pretty, but not exactly penury.

At this juncture in US history, however, there are worse things that could happen than a one-off, 8% drop in grant funding. The nation might, for example, continue to slip and fudge into inexorable debt and decline — a bad thing for scientists as well as the public. Going over the cliff would avert that: even the new apparatchiks on the Central Committee of the Chinese Communist Party might wake up on 1 January, blink and think: good God, perhaps America isn’t finished, after all.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/vlzvwz

As FASEB and other science supporters know, research and development spending has not and will never veer far from its historical level of one-seventh of US discretionary spending. If taxes were raised and defence spending cut, the long-term outlook for non-defence discretionary spending would brighten considerably.

After all, the ‘cliff’ isn’t a cliff at all. It is simply a new baseline, with proper taxes paid, spending reduced and the poor protected. Once it is set, the path may be open for selective spending boosts — including, perhaps, in research — as well as tax reductions. That’s why people such as Peter Orszag, a member of the Institute of Medicine and President Barack Obama’s first budget director, and anti-austerity economist Paul Krugman say that going over the cliff may be the best path to a reasonable budget settlement.

Concern that the cold bath of spending cuts and tax rises will send the economy into recession is legitimate. But economists do not actually know how fiscal tightening affects economic growth. Just last month, the International Monetary Fund revised its estimate of the fiscal multiplier — the dollars of economic activity generated by each dollar of government spending — from 0.5 to “in the range of 0.9 to 1.7”, admitting, really, that it can’t read the complex relationship between fiscal tightening and economic growth.

Sure, Obama and his lieutenants need to say publicly that the nation must avoid going over the cliff. In conducting negotiations for a deficit reduction that does not savage public spending, however, his willingness to take the drop is his single most powerful weapon.

Democratic politicians such as Obama increasingly see scientists as part of their constituency (in the final tracking poll by the *Washington Post* and the ABC, Obama beat Romney 60% to 38% among voters with a postgraduate degree). So the bleating about the cliff from scientific societies merely serves to lessen that resolve. The science lobby, in other words, is pushing the president to fold.

But the president shouldn’t give an inch, and, if need be, he should be ready to jump off that cliff. As he smiles for the television cameras and joshes with House Speaker John Boehner, I hope that Obama will quote an old enemy to his new friend. “Go ahead,” should be his whispered message, “make my day.” ■

Colin Macilwain *writes about science policy from Edinburgh, UK.*
e-mail: cfmworldview@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

BOTANY

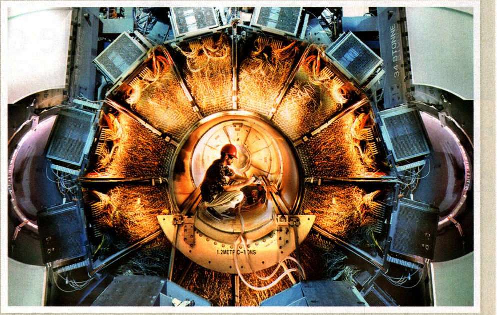
Plant fertilization protein found

Fertilization in flowering plants is dependent on a protein that is secreted by the egg cell and activates incoming sperm.

Stefanie Sprunck at the University of Regensburg in Germany and her colleagues show that, in the model plant *Arabidopsis thaliana,* the arrival of sperm cells near the egg causes the release of a protein they call EGG CELL 1 (EC1). This triggers the redistribution of a second protein — one linked to fusion of the sex cells, or gametes — from inside the sperm to the sperm cell surface.

Sperm cells interacting with mutant *Arabidopsis* eggs that have faulty *ec1* genes failed to fuse, and the plant's pollen tubes continued to deliver sperm into the embryo sac. These results suggest that EC1 controls gamete fusion.

Science **338, 1093–1097 (2012)**



SLAC NATL ACCELERATOR LAB.

PARTICLE PHYSICS

Time's arrow in B mesons

A cornerstone of theoretical particle physics — the idea that not all processes run in the same way forwards in time as they do backwards — has been observed directly for the first time.

Members of the BaBar Collaboration trawled data from their experiment (pictured), which ran at the SLAC National Accelerator Laboratory in Menlo Park, California, from 1999 to 2008. The researchers identified B-meson decay chains that were time reversals of each other, and a comparison of the decay rates revealed a strong asymmetry. Earlier experiments have caught hints of time-reversal violation but failed to distinguish it clearly from violations of other fundamental symmetries.

Phys. Rev. Lett. **109, 211801 (2012)**

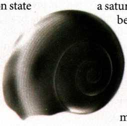
For a longer story on this research, see go.nature.com/258vei

CLIMATE CHANGE

Carbon drop in snail shell shock

Free-swimming snails show shell damage in water conditions that are likely to become more prevalent as the climate warms. By 2050, the top 200 metres of the Southern Ocean are likely to become under-saturated in a form of calcium carbonate called aragonite — a component of many shells. If aragonite structures are placed in waters in which the saturation state is less than one, they gradually dissolve.

Geraint Tarling of the British Antarctic Survey in Cambridge, UK, and his team analysed the shells of *Limacina helicina antarctica* pteropods (**pictured**) captured from the top 200 metres in an upwelling region of the Southern Ocean in 2008. Their shells were thinner and more porous than those captured elsewhere. In the laboratory, eight days of immersion in waters with a saturation state of between 0.94 and 1.12 produced similar levels of damage. Aragonite-shelled animals, important to food and carbon cycles, may decline sooner than expected in the Southern Ocean, the authors say.



NINA BEDNARŠEK/BERTRAND LÉZÉ

Nature Geosci. **http://dx.doi.org/10.1038/ngeo1635 (2012)**

ZOOLOGY

Blue whales roll with it

Blue whales (*Balaenoptera musculus*) can perform 360° rolls, an impressive manoeuvre for the largest animal ever to have lived.

Jeremy Goldbogen of the Cascadia Research Collective in Olympia, Washington, and his colleagues tagged blue whales off the coast of California with sensors that provided information regarding the animals' speed, depth and body orientation. Of 22 whales tagged for an average of 6.7 hours, 11 were recorded performing rolls. In total, 44 rolls were observed, all during foraging.

The authors suggest that these rolls serve a dual purpose, allowing the animal both to re-orient its body to capture the maximum amount of krill prey, and to better visualize the prey and its surroundings.

Biol. Lett. **http://dx.doi.org/10.1098/rsbl.2012.0986 (2012)**

CHEMISTRY

Instant steam from sunlight

Nanoparticles that efficiently absorb light energy and convert it into heat can act as miniature steam generators in a liquid.

Naomi Halas, Peter Nordlander and their colleagues at Rice University in Houston, Texas, used lenses to focus sunlight on carbon or gold–silica nanoparticles suspended in water. Within a few seconds, steam at a temperature well above 100 °C was generated around the particle surfaces and bubbled away, whereas the bulk of the water heated up only gradually.

This method of using sunlight to generate high-temperature steam could be used to sterilize waste or surgical instruments without the need to boil a large volume of fluid, the authors suggest. The same effect may improve distillation: sunlight focused on nanoparticles in an ethanol–water mixture produced vapours richer in ethanol than conventional thermal distillation.

ACS Nano **http://dx.doi.org/10.1021/nn304948h (2012)**

PALAEONTOLOGY

Fossil hints at star's salty past

Modern echinoderms —invertebrates such as brittlestars and sea urchins — live only in open seas, but fossils from Europe suggest that this has not always been the case.

A team led by Mariusz Salamon of the University of Silesia, Poland, examined fossils of the *Aspiduriella similis* brittlestar from a quarry in southern Poland. The fossils were embedded in limestone dated to the Middle Triassic period, more than 240 million years ago. Minerals and geological structures within the rocks suggest that the fossils formed in conditions with very high salt levels, such as those present in hypersaline coastal waters. Few other fossils were found in the rock, also pointing to harsh living conditions. Echinoderm fossils have been used as indicators of open marine environments, something the authors caution against.

PLoS ONE 7, **e49798 (2012)**

BIOLOGY

Switching off cancer resistance

Modern therapies can target specific pathways in cancer cells, but the cells often become drug resistant. René Bernards of the Netherlands Cancer Institute in Amsterdam and his colleagues have identified a gene involved in resistance, and have found a way to stop it in its tracks.

Resistance can be caused by mutations in genes or proteins that are not directly targeted by a drug. Bernards' team used a genetic technique called RNA interference to investigate the effects of shutting down thousands of human genes.

The researchers found that when the gene *MED12* was switched off, the cells in a variety of cancers became resistant to a range of anti-cancer drugs. Suppressing this gene activates the transforming growth factor β receptor (TGF-βR) and, conversely, inhibiting the signalling through this receptor in drug-resistant cells eliminates the resistance.

Cell 151, **937–950 (2012)**

OPTICS

Technology for thinner probes

A single optical fibre could form the basis of thinner endoscopes — long imaging probes with medical, military and industrial uses.

Current endoscopes are made up of millimetre-sized bundles of up to 100,000 fibres. Each fibre transports a single mode of light wave coming from the object being imaged, because the mixing of modes can cause light-wave distortion. Wonshik Choi at Korea University in Seoul and his colleagues have used a single 200-micrometre-wide fibre to transport multiple modes, by measuring and reverse engineering the distortion that each mode suffers. The authors used their technique to make a three-dimensional map of a sample of rat intestine.

Phys. Rev. Lett. 109, **203901 (2012)**

COMMUNITY CHOICE

The most viewed papers in science

STEM CELLS

Cancer-drug infertility reversed

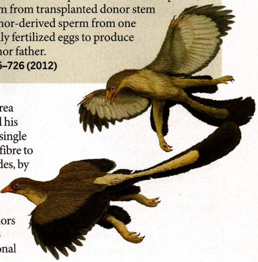
HIGHLY READ on www.cell.com in November

Transplanted testicular stem cells can restore fertility to macaque monkeys made infertile by chemotherapy.

Kyle Orwig at the University of Pittsburgh in Pennsylvania and his colleagues took stem cells capable of developing into sperm from macaques and marked them with a lentivirus. After inducing infertility in the 12 adult and 5 prepubescent donors, the researchers returned the marked cells to the animals. Marked genetic material later appeared in the sperm of nine of the adults, and three of the juveniles when they reached maturity.

The team then ran a similar experiment, transplanting cells from the donors into other macaques. Of six adult recipients, two produced sperm from transplanted donor stem cells. Moreover, donor-derived sperm from one recipient successfully fertilized eggs to produce embryos with a donor father.

Cell Stem Cell 11, **715–726 (2012)**



CARL BUELL/J. VINTHER/BRISTOL UNIV.

PALAEONTOLOGY

Birds of a different feather

The wings of ancient birds and feathered dinosaurs that lived some 150 million years ago may have been less like those of modern birds than previously thought.

Contemporary birds share a common wing design, with a single feather layer. But Nicholas Longrich of Yale University in New Haven, Connecticut, and his colleagues identify separate layers in their fossil analyses of the wings of the Jurassic bird *Archaeopteryx lithographica* (**pictured**) and the Cretaceous feathered dinosaur *Anchiornis huxleyi*.

The slender feather shafts found in these prehistoric plumages would make the feathers weak by modern standards but, when stacked, may have formed a structure strong enough to generate lift. However, the layers would have limited the airflow through the wing, which is used by modern birds for take-off and low-speed flight, so these prehistoric flyers probably glided or parachuted down from trees, the authors say.

Curr. Biol. **http://dx.doi.org/10.1016/j.cub.2012.09.052 (2012)**

➲ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

EU budget limbo

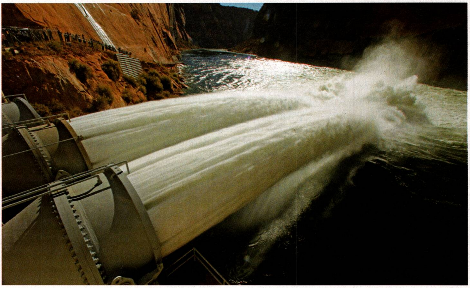
Talks between European heads of state ended on 23 November without any agreement being reached on the European Union (EU) budget for 2014–20. Negotiations will continue next year, but the EU Horizon 2020 research programme is facing cuts of 12% in a deal proposed by Herman Van Rompuy, president of the European Council, who is chairing the negotiations. See page 638 for more.

Space deal

Europe's research ministers have agreed a budget of €10.1 billion (US$13 billion) for the European Space Agency (ESA) for the next few years, and have set out the next steps for replacing the Ariane 5 satellite-launching rocket. The settlement means that ESA's science programme will have flat funding of about €500 million per year between 2013 and 2017. See page 645 and go.nature.com/dinqbk for more.

UK energy deal

Britain's energy and finance ministries have struck a deal to support financing for low-carbon energy. On 23 November, the government said that utility firms could triple customer charges that support nuclear, wind, solar and other low-carbon electricity sources, bringing such funding to £7.6 billion (US$12.2 billion) annually (in real terms) by 2020. The United Kingdom wants 30% of its electricity to come from renewables by then, up from 11% today. But the deal — which foreshadowed an energy bill released this week — saw politicians drop a proposal to eliminate carbon emissions from the electricity sector almost entirely by 2030. See go.nature.com/ly4mhe for more.



B. STRONG/REUTERS

Grand Canyon flooded to restore habitats

A gush of water into the Colorado River marked the start of efforts to rebuild beaches and sandbars by redistributing sediment along the Grand Canyon. Water released from the Glen Canyon Dam flowed fastest between 9 p.m. on 19 November and 10 p.m. the next day, with flows reaching around 1,200 cubic metres per second, according to the US Department of the Interior. The project, which follows 16 years of research on the potential effects of the releases, aims to restore vegetation and boost populations of fish and other wildlife.

Progress on HIV

In the past two years, the number of people accessing antiretroviral therapy for HIV has increased by 63% globally. Moreover, in the past six years AIDS-related deaths have fallen by one-quarter (to around 1.7 million in 2011). And in the past decade the number of new infections has fallen by 20% (to around 2.5 million in 2011) — with declines of more than 50% in 25 low- and middle-income countries, many of them in Africa. The encouraging statistics were reported on 20 November by the Joint United Nations Programme on HIV/AIDS.

Polio setback

The Global Polio Eradication Initiative will miss its goal of stopping spread of the viral disease this year, says a report released this week by the programme's Independent Monitoring Board. Of the four countries where the disease still exists, three — Afghanistan, Pakistan and Chad — are making progress against polio. However, the fourth, Nigeria, has seen cases double between 2011 and 2012, accounting for more than half of the 175 cases recorded globally this year.

UK science advice

The UK Parliamentary Office of Science and Technology has been spared heavy budget cuts. The office, which provides politicians with analysis of scientific issues, was facing cuts of up to £98,000 (US$157,000), or 17% of its budget. But after a protest backed by two former British science ministers, the office has secured financing until April 2015. See go.nature.com/b3gph9 for more.

BUSINESS

Egg-free flu jab

A seasonal influenza vaccine made without growing the virus in fertilized chicken eggs was approved for the first time by the US Food and Drug Administration on 20 November. Flucelvax uses flu strains grown in animal cells and is made by drug giant Novartis, headquartered in Basel, Switzerland. Like other flu vaccines, it protects against the three strains of flu predicted to be the most prevalent during the coming flu season. See go.nature.com/1nmiqc for more.

CHUCK NACKE/TIME LIFE PICTURES/GETTY

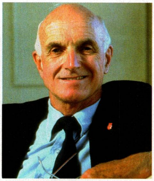
PEOPLE

Ape expert cleared

Sue Savage-Rumbaugh, former executive director of the Iowa Primate Learning Sanctuary in Des Moines, is to return to the centre after an investigation committee found no evidence to support allegations that she had failed to care for bonobos there. In a statement released on 20 November, the sanctuary said that the committee had "encountered significant counter-evidence against the claims". Savage-Rumbaugh had been placed on leave in September pending the results of the enquiry. See go.nature.com/caqgik for more.

Fraudster punished

Nutrition researcher Eric Smart has agreed to refrain from making grant applications to the National Institutes of Health for seven years. The US Office of Research Integrity reported on 20 November that Smart fabricated figures in 10 papers and 7 grant applications over 10 years while he was at the University of Kentucky in Lexington. See go.nature.com/h4srje for more.



Transplant pioneer

Joseph Murray (**pictured**), who performed the first successful human organ transplant, died on 26 November, aged 93. In 1954, he transplanted a kidney between identical twins. Murray, a director of the Surgical Research Laboratory at Harvard Medical School in Boston, Massachusetts, shared the 1990 Nobel Prize in Physiology or Medicine.

Illegal profits

Sidney Gilman, a neurologist at the University of Michigan in Ann Arbor, has been charged with insider trading that netted two hedge-fund companies a total of US$276 million. Gilman gave hedge-fund manager Mathew Martoma of CR Intrinsic Investors, based in Stamford, Connecticut, early news of safety data from clinical trials of an experimental drug for Alzheimer's disease. Gilman has agreed to pay a settlement of more than $234,000, but Martoma faces criminal charges. See go.nature.com/n8tnbi for more.

EU health chief

Conservative Maltese politician Tonio Borg was approved as Europe's commissioner for health and consumer affairs, after the European Parliament voted in favour of his appointment on 21 November. Critics fear that the devout Catholic could attempt to derail European Union funding for human embryonic stem-cell research. Borg replaces John Dalli, a Maltese politician who resigned on 16 October as a result of corruption charges.

RESEARCH

Emissions peak

The concentration of carbon dioxide in the atmosphere hit a record high of 390.9 parts per million in 2011, the World Meteorological Organization reported on 20 November. Levels of the potent greenhouse gases methane and nitrous oxide also reached new highs last year. See go.nature.com/8hziue for more.

Contagion concerns

Four more cases of an infection caused by a novel coronavirus — the viral family behind severe acute respiratory syndrome (SARS) — were reported by the World Health Organization on 23 November, bringing the tally to six. Three cases, including one fatality, occurred in Saudi Arabia; the fourth was in Qatar. Two of the new cases are from one household, raising the possibility that the virus can be transmitted between people, not just by contact with infected animals. See go.nature.com/9f5zwq for more.

Flight ban flouted

A beagle breeder has dodged an airline's ban on transporting animals bound for research labs by stating that the dogs would not be harmed. The puppies were bound for Advinus, a contract-research organization in Bangalore, India, that uses beagles in drug-toxicity work in which the animals are euthanized. But the breeder, Beijing Marshall Biotechnology, a branch of Marshall BioResources of North Rose, New York, told Cathay Pacific airline that the dogs would not be hurt or killed. The New York firm said last week that it is investigating the incident. See go.nature.com/zurgnl for more.

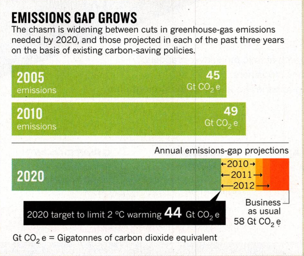
COMING UP

3–7 DECEMBER
Preliminary findings from James Cameron's dive to the bottom of the Mariana Trench, and new results from NASA's Curiosity rover on Mars, are discussed at the American Geophysical Union's meeting in San Francisco, California.
fallmeeting.agu.org/2012

5 DECEMBER
UK scientists begin to drill into Antarctica's subglacial Lake Ellsworth, buried under more than 3 kilometres of ice (see *Nature* **491,** 506–507; 2012).
www.ellsworth.org.uk

TREND WATCH

As climate negotiators gather in Doha, the United Nations Environment Programme has released a report warning that pledges to cut greenhouse-gas emissions fall short of what is needed to have a "likely" (greater than 66%) chance of limiting global temperature rise to 2 °C. By 2020, annual emissions must be no more than 44 gigatonnes (Gt) of carbon dioxide and equivalent gases — but even the strictest pledges fall 8 Gt short of that target, a gap that has grown since last year.



EMISSIONS GAP GROWS

The chasm is widening between cuts in greenhouse-gas emissions needed by 2020, and those projected in each of the past three years on the basis of existing carbon-saving policies.

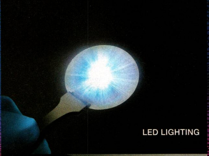
Gt CO_2 e = Gigatonnes of carbon dioxide equivalent

SOURCE: UNEP

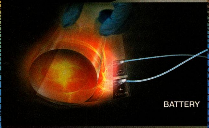
NATURE.COM
For daily news updates see:
www.nature.com/news

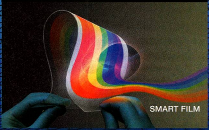
BLUESTONE
GLOBAL TECH

www.bluestonegt.com



TOUCH PANEL





Redefining Everything™

Bluestone Global Tech (BGT) is the first company to provide high quality, large-area graphene films—over 50".

BGT is using this breakthrough material to enhance the performance of everyday technologies including LED lighting, flexible electronics, batteries and other nanoelectronics.

What innovation is possible with atomic monolayer graphene?
Your search for graphene films stops here.
Order your superior quality graphene film at **www.bluestonegt.com.**

GRATOM-A™
Large-area graphene films up to 50"

GRATOM-R™
Low sheet resistance and high optical transmittance graphene film

GRATOM-M™
High carrier mobility monolayer graphene film

GRATOM-S™
Isolated monolayer graphene single crystals with large domain size

1-855-276-8290 • www.bluestonegt.com • info@bluestonegt.com

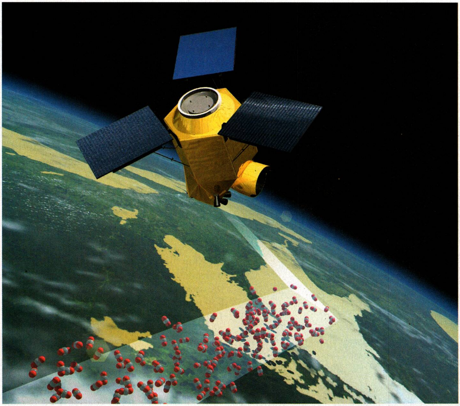
NEWS IN FOCUS

BIOFUELS How Brazil's ambitious ethanol drive has sputtered **p.646**

EASTERN EUROPE Charges of bad science and cronyism roil Bulgaria **p.649**

EARTH OBSERVATION A cut-rate scheme for monitoring the atmosphere **p.650**

AFTER KYOTO The treaty expires with emissions skyrocketing **p.653**



ASTRIUM

ESA's funding shortfall is bad news for CarbonSat, a mission aiming to track atmospheric carbon dioxide.

FUNDING

Space budget blow to climate science

Economic difficulties take their toll on European Space Agency's Earth-observation programme.

BY EDWIN CARTLIDGE

For Europe's space chiefs, the outcome of last week's European Space Agency (ESA) budget negotiations was better than expected, given the continent's economic troubles. But for Volker Liebig, ESA's head of Earth observation, there is a sting in the agreement. The multi-year budget that member states approved — which falls some €2 billion (US$2.6 billion) short of ESA's proposed spending of about €12 billion — could force him to postpone or cancel a mission aimed at pinning down the mysterious carbon sinks that are slowing the rise of greenhouse gases in Earth's atmosphere.

Ahead of the budget negotiations in Naples, Italy, on 20–21 November, Liebig had hoped to secure around €1.25 billion for new research satellites. With France, Italy and Spain contributing much less than expected, he received €1.9 billion for Earth-observation projects. But €808 million has already been allocated for a new generation of weather-forecasting satellites, leaving him with little more than €1 billion for research missions. "We have to discuss with scientists in the next few weeks what to do," Liebig says. "But we will not be able to develop all the science satellites we wanted to."

Most vulnerable, he says, is a planned €250-million climate-change mission scheduled for launch in about 2018. One of the two contenders for the mission, CarbonSat, would map atmospheric concentrations of carbon dioxide and methane at high-enough resolution to investigate a long-standing puzzle: why only about half of the CO_2 emitted by human activities remains in the atmosphere. Scientists assume that the rest is absorbed largely by the oceans and plants, but ground-based monitoring stations are too few and far apart to pinpoint the sinks.

Satellites could fill in the gaps in the picture, but in April ESA lost contact with Envisat, the one satellite providing such data (see *Nature* **484,** 423–424; 2012). Neither Japan's existing Greenhouse Gases Observing Satellite nor NASA's Orbiting Carbon Observatory-2 (OCO-2), scheduled for launch in 2014, will map greenhouse-gas concentrations in as much detail as CarbonSat, which would survey the whole globe with a resolution of 4 square kilometres. "The information that it would collect is essential for developing, implementing, and monitoring greenhouse-gas-emission policies," says atmospheric physicist David Crisp of NASA's Jet Propulsion Laboratory in Pasadena, California, who is the science team leader of OCO-2. "A timely launch of this satellite should be among the highest priorities of ESA".

CarbonSat's competitor for ESA funding, FLEX, would also help to pin down carbon sinks, by measuring the faint fluorescence generated by plants during photosynthesis — a measure of how efficiently they absorb carbon. "The last thing we want to do is to destroy the forests or whatever is absorbing almost half of the CO_2 that we are emitting," says Crisp. "Wouldn't it be good to know where these ▶

➲ **NATURE.COM**
For more from the ESA ministerial meeting, see:
go.nature.com/dinqbk

processes are occurring?"

However, there was better news for other ESA programmes. Europe's Ariane 5 rocket launcher, which is more expensive than competitors, was the focus of fraught discussions: Germany argued for a more powerful and versatile upgrade, whereas France maintained that it would be better to switch straight to a new and more economical launcher. Following late-night discussions, ministers decided to fund both designs over the next couple of years and to review progress in 2014.

They also reached a deal on how to pay for Europe's contribution to operating the International Space Station between 2017 and 2020. The costs will be covered in kind by a German-backed plan to provide the propulsion and avionics for NASA's Orion manned spacecraft. ESA also agreed to Russian involvement in its twin ExoMars missions, an ambitious programme of orbiters and landers scheduled for launch in 2016 and 2018. NASA pulled out of the project earlier this year.

But ESA's science programme faces a squeeze: it will receive €508 million a year for the five-year period from 2013 to 2017. Although slightly higher than its 2012 funding of €480 million, thanks to the contributions from new member states Poland and Romania, after inflation is taken into account this effectively amounts to a cut. Willy Benz of the University of Bern, chair of ESA's Space Science Advisory Committee, says that this could force the agency to delay a future large mission; cancel mission extensions for existing probes; or cancel smaller missions.

Benz thinks that the science programme got less than expected because in hard economic times spending is channelled towards activities that can more directly boost industry, such as designing and building new launchers.

"I would say that the budget outcome was the best we could have hoped for given the economic circumstances," says Benz. "But if you cut budgets in the science programme, you cut science. There is only so much you can save by reducing travel or not making phone calls." ■

Brazil has struggled to sustain its production of biofuel from sugar cane.

RICKEY ROGERS/REUTERS

ENERGY

Growth of ethanol fuel stalls in Brazil

Shortages are a sobering lesson from a biofuels pioneer.

BY CLAUDIO ANGELO IN BRASÍLIA

"A new moment for mankind." That was how Brazil's former president, Luiz Inácio Lula da Silva, described his country's biofuel boom in March 2007. Back then, Brazil was the poster child of ethanol fuel, its output second only to that of the United States. Fermenting the sugars in the country's abundant sugar cane produced a motor fuel that lowered carbon dioxide emissions, and many saw Brazil as a model for how the world could shed its addiction to oil, creating jobs along the way.

Five years on, Lula's vision has tarnished. Biofuels are falling from grace around the world as critics charge that devoting millions of hectares of agricultural land to fuel crops is driving up food prices and that the climate benefits of biofuels are modest at best. But the fall has been hardest in Brazil, where government policies have compounded the effects of the global economic downturn.

Domestic consumption of liquid ethanol this year has been 26% lower than for the same period in 2008. Forty-one of the country's roughly 400 sugar-cane ethanol plants have closed over that time. The price of pure ethanol at the pump is so high that in most states it is cheaper to fill up flexible-fuel cars with petrol blends that contain about 20% ethanol. The shift back to fossil fuels, combined with rapid growth in the number of cars on the roads (see 'Fuelling Brazil's transport boom'), has worsened city smog and caused emissions in the transport sector to spike at about 170 million tonnes of CO_2 in 2011, up from less than 140 million tonnes in 2008. "We are increasing the world's GDP: we are buying more oil and spending more on pollution-related health care," jokes Ildo Sauer, who studies energy policy at the University of São Paulo and is a former director of the state oil giant Petrobras.

Brazil's ethanol roller coaster is a sobering example of what can happen when climate and energy planning clash with economic decision-making. It began with the 2008 economic crisis, which staunched new investments in the sector just as it was expanding rapidly, and deep in debt. Rather than developing new plantations, the industry fell back on harvesting cane from older, less-productive sites, and average yields plummeted from 115 tonnes per hectare in 2008 to 69 tonnes this year. Combined with two bad harvests, this has forced Brazil to import some 1.5 billion litres of maize (corn) ethanol from the United States over the past 2 years.

But the killer blow came when the government decided to freeze the price of petrol and diesel to keep inflation under control, leaving biofuels less competitive. On the very night that current President Dilma Rousseff gave the closing speech of the Rio +20 conference in June — the final agreement of which promised to phase out fossil-fuel subsidies — the government said it would be reducing a federal fuel tax to zero. "We have taken away jobs from agroindustry, stalled growth and worsened the air of our cities for the sake of inflation control," says Luiz Horta, a bioenergy researcher at the Federal University of Itajubá.

Meanwhile, the government has tried to stimulate the economy with tax breaks on the sale of new cars. That, combined with the cost of pure ethanol, has meant that "the share of alcohol in our transport fuel matrix has dropped from 55% in 2008 to 35%", says André Ferreira, head of the Institute for Energy and the Environment, a think-tank in São Paulo.

According to Antônio de Pádua Rodrigues, technical director and acting president of UNICA, Brazil's sugar-cane industry association, the government knows that the situation is unsustainable. It has promised the industry that petrol prices will go up next year, and that the blend of ethanol will rise from 20% to 25%, the maximum allowed by law. But it will take time for the industry to bounce back from its poor fortune, and ethanol is likely to remain scarce and expensive for the next two years, say Rodrigues and Horta.

Now, Brazil hopes to tap into a new biofuel source: second-generation ethanol, produced from the tough cellulose in plant stalks. Cellulose is difficult to break down and ferment, but several facilities in the United States are on the verge of making commercial cellulosic ethanol — for example, by using specialist enzymes to break down the long-chain cellulose molecules — and Brazil doesn't want to be left behind.

➲ NATURE.COM
Read more in *Nature*'s chemistry and energy Outlook.
go.nature.com/q6zodw

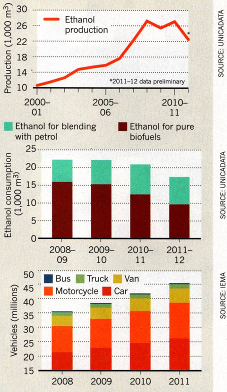
In December last year, the Brazilian Development Bank launched a 1-billion-real (US$481-million) credit line to stimulate research and development in cellulosic biofuels and other advanced sugar-cane technologies. The Center for Sugarcane Technology, an industry-sponsored organization based in São Paulo, has taken up a 357-million-real loan to build a cellulosic ethanol plant next year, which would use waste plant matter from conventional sugar-cane fermentation. "We can double fuel yield per hectare when the technology is mature", says Oswaldo Godoy, a project manager at the organization.

The Brazilian Agricultural Research Company (EMBRAPA) is also throwing its weight behind bioenergy. Its president, Maurício Lopes, a geneticist who took office in October, has promised to build up research on biomass technology and double EMBRAPA's funding for that area, which today stands at a modest 24 million real per year. "I want to believe that the current state of the ethanol sector is a temporary blip," he says. Lopes says that Brazil will be "unbeatable" once cellulosic technology matures. "No other country has the logistics we have in place, or the number of different species we can derive ethanol from."

But cellulosic ethanol won't be a quick fix, says Horta. "Nothing shall compete with conventional sugar-cane ethanol until 2050." ■

FUELLING BRAZIL'S TRANSPORT BOOM

More vehicles and falling ethanol production is prompting a switch to petrol blends.



SOURCE: UNICADATA

SOURCE: UNICADATA

SOURCE: IEMA

EPIDEMIOLOGY

Daily dose of toxics to be tracked

Exposome studies will tie environmental exposure to biological triggers of disease.

BY EWEN CALLAWAY

Think of it as a benevolent Big Brother. European researchers are gearing up to monitor thousands of people by giving them smartphones to record the chemicals to which they are exposed every day.

Two projects this week announced that they had won a combined €17.3 million (US$22.4 million) from the European Commission to study the 'exposome' — the effects of environmental exposures on health. The researchers hope that the four-year studies will benefit public health in ways that genome research so far has not.

Genome-wide association studies, in which scientists search for genetic variants linked to disease, have failed to fully explain why some people are more susceptible than others to chronic diseases, such as type 2 diabetes. "There's been too much emphasis on genetic factors, which contribute relatively little to disease compared with environmental factors," says Martyn Smith, a toxicologist at the University of California, Berkeley, who is participating in the newly funded Exposomics project. Paolo Vineis, an environmental epidemiologist at Imperial College London, leads the €8.7-million project.

Subjects will carry smartphones equipped with sensors to measure exposures, and their blood will be analysed to monitor molecular changes. Most participants are already involved in other long-term health studies. One goal is to look for biomarker differences between people walking through areas with low air pollution and those exposed to urban fumes, in order to understand the triggers for conditions such as heart disease, asthma and lung cancer.

Vineis's exposomics approach has already uncovered gene-expression signatures that link people's leukaemia risk with their exposure to heavy metals and other toxic chemicals, for example.

The second project, the €8.6-million Human Early-Life Exposome, will focus on children and pregnant women. Children are more susceptible to environmental influences because their bodies are smaller and their organs are still developing, says epidemiologist Martine Vrijheid at the Centre for Research in Environmental Epidemiology in Barcelona, Spain, who heads the project. The researchers will track disease biomarkers to assess the effects of environmental exposures on growth, obesity, immune development and asthma. Both projects will generate vast amounts of data, and Vineis and Vrijheid are developing data-sharing policies to enable other researchers to mine the resource.

Interest in exposomics is also growing in the United States. This year, the US National Research Council called for greater investment in exposome research, and the National Institute for Environmental Health Sciences plans to make it a priority, although it has yet to invest in any projects as large as the European efforts, says the institute's David Balshaw. "We see this as a major priority," he says. ■

ARE YOU AS AMBITIOUS AS WE ARE?

The University of Exeter, a member of the Russell Group and firmly positioned in the UK Top 10 and the global Top 200, is embarking on a major expansion of its Medical School in both clinically-related and basic science research as part of an ambitious growth plan.

BASIC SCIENCE

In basic science, the Medical School is part of a multi-million pound interdisciplinary initiative with other Colleges of the University to investigate cellular dynamics, to develop a quantitative understanding of cell behaviour and the underlying biochemical and biophysical properties of cells and to determine how defects in cellular dynamics cause disease.

We are seeking to appoint world-leading researchers in the areas of Molecular Cell Biology, Genetics, Epigenetics, Systems Medicine, preferably with interests related to neuroscience, diabetes, vascular biology, ageing, and/or environmental challenge.

Either non-clinical or clinically qualified individuals may apply for the following basic science posts:

- Professors of Functional Cell Biology
- Professors of Systems Medicine
- Professors of Genetics/Epigenetics

CLINICAL AND HEALTH SERVICES RESEARCH

In clinical and health services research, the Medical School's ambitions focus on improving patient care, patient stratification, and health and wellbeing of the population. There are highly successful collaborations with health researchers in other departments of the University, particularly the world-leading Mood Disorders Centre, the Business School and in exercise science.

We are seeking to appoint world-leading researchers with expertise in clinical trials, health services or applied health research, experimental medicine investigating the mechanisms underlying disease or responses to treatments, diagnostics, innovative approaches to health data usage, systems medicine, or behavioural change.

The clinical research posts available are:

- Professors of Clinical or Health Research or Clinical Trials
- Professors of Complex Health Interventions
- Professors of Experimental Medicine
- The Gillings Chair in Applied Health Statistics
- Professors of Diagnostics

More detail on the School's research strategy and the posts can be found at **www.exeter.ac.uk/medicine/posts**

UNIVERSITY OF EXETER | MEDICAL SCHOOL

THE SUNDAY TIMES
UNIVERSITY OF THE YEAR 2012-13
UNIVERSITY OF EXETER

U231227R

POLICY

Funding protest hits Bulgarian research agency

Petition alleges that grant competition was mismanaged and cash channelled to bad science.

BY ALISON ABBOTT

Bulgarian scientists have never had much faith in their research ministry, but the outcome of this year's grant competition has provoked a unprecedented storm of outrage.

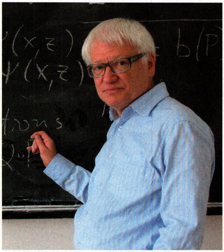
On 22 November, a detailed front-page report in the national newspaper *SEGA* presented a hair-raising list of allegations, ranging from large funding allocations to companies and foundations with no experience in scientific research, to alleged conflicts of interest involving geologist Rangel Gjurov, who chairs the executive board of the Ministry of Education and Science's Bulgarian National Science Fund (BNSF). Gjurov advocates an unproven 'corkscrew' theory of earthquake prediction.

A growing number of scientists are now alleging that the funding agency is funnelling research cash towards bad science, and unfairly favouring those with close ties to the agency. This week, more than 400 researchers sent a petition to the prime minister and key science policy-makers, demanding a reassessment of the competition. "We don't have any idea why some projects didn't win, because we had no feedback from reviews," says materials scientist Victor Atanasov, from the University of Sofia, who helped to organize the petition and whose application for graphene research was rejected. "And we have no idea why some projects won because no rankings were provided."

The protest comes at a time when support for science is at a low ebb. Since 2009, university budgets have fallen by more than 20% and the budget for the Bulgarian Academy of Sciences, which runs 41 research institutes in the country, has dropped by more than 40%. "It is obviously important that the small amount of money we have is spent appropriately," says Atanasov.

The disputed competition was announced in May by the BNSF, offering grants totalling 14.8 million leva (US$9.8 million). Projects within six priority areas — including material sciences and biotechnology — were invited to bid for up to half a million leva each, and 95 were eventually selected from roughly 1,200 applications.

The petition says that two of the projects, awarded 678,000 leva between them, were supported only because of their connection with Gjurov. Many geologists are sceptical of Gjurov's earthquake-prediction theory, which links geological processes, phases of the Moon and climate change. Some of the winning projects involve his recently qualified PhD student and a fellow advocate of the corkscrew theory. Gjurov did not respond to *Nature*'s questions about the allegations against him.



Emil Horozov quit as head of Bulgaria's National Science Fund over concerns about corruption.

MITKO YANCHEV

SATELLITE CONCERNS

The petition also raises concerns about a project to build a nanosatellite, which snagged 480,000 leva for two companies, called Bulcube and Space Research. Both were registered as new companies only on 16 July, the deadline for the call, and each declared just 50 leva in capital. The project's principal investigator, Valery Golev, head of astronomy at the University of Sofia, says that the satellite project will provide useful training to help Bulgaria develop its space activities, and that he believes the new companies to be competent.

In a written statement to *Nature*, a spokesperson for the ministry said that the large number of applications in the competition demonstrated that scientists trusted the BNSF. The ministry did not respond to the allegations that grants were awarded inappropriately, or explain their choice of grant-winners.

The protest comes as former BNSF director Emil Horozov, who alleged corrupt practices in research funding two years ago, finds himself the target of minor mismanagement charges. Horozov, a mathematician, became BNSF director in January 2010 with a mandate to reform the agency. He and his team produced a report detailing irregularities involving more than half of the projects funded in 2008 and 2009, which Horozov passed on to ministry officials in November 2010. But Horozov resigned in February 2011, believing that the report was being ignored (see *Nature* **472**, 19; 2011). "We are talking here of tens of millions of euros, perhaps even hundreds of millions," claims Horozov. "I want to make sure the bad practices are exposed and stopped."

A year later, the Public Financial Inspection Agency (ADFI) in Sofia — which is investigating the allegations — told Horozov that its inquiry would include his tenure as director. The ADFI subsequently identified 30 procedural misdemeanours, and blamed nine minor infractions on Horozov.

Horozov denies any impropriety, and alleges that the charges are a smokescreen to cover up the ADFI's failure to address much more serious problems raised in his report. The ADFI says that its investigation followed proper procedures, and its findings are publicly available.

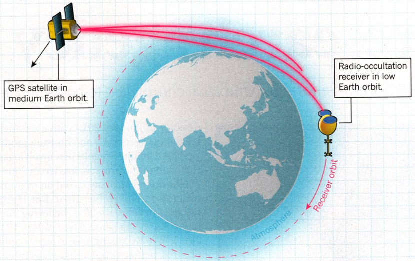
Horozov's report alleged, for example, that roughly 2.4 million leva were channelled without proper review to projects involving four foundations, which offered logistical services for research into Black Sea resources and marine infrastructure. One of the projects was actually rejected in 2009, but then approved in a closed meeting of the BNSF's executive board in 2011, which also threw in a top-up donation of 800,000 leva to another project. There has been no formal review of the projects' progress.

All four foundations are registered in the name of businessman Galin Dimitrov. He told *Nature* that he is president of the Foundation for Development and Implementation of Public Resources, but is only a donor to the others, adding that the foundations offer legal and finance expertise to scientists. His goal, he says, is to make "Bulgaria a better place to live".

Scientists, however, believe that Bulgaria's meagre funds are being channelled away from worthy research. "We don't believe that the reviewing process was fair," says microbiologist Pavlina Dolashka from the Bulgarian Academy of Science's Institute of Organic Chemistry in Sofia, whose application on vaccine development was rejected. "As some of the grants were very large, we need to protest." ■

BENDING FOR DATA

Radio signals from Global Positioning System (GPS) satellites bend as they pass through the atmosphere; the amount of bending relates to the atmospheric temperature and moisture levels. Each of these 'occultation events', picked up by a receiver, results in hundreds of data points at different elevations in the air column.


GPS satellite in medium Earth orbit.
Radio-occultation receiver in low Earth orbit.
Receiver orbit
Atmosphere

ATMOSPHERIC SCIENCE

Microsatellites aim to fill weather-data gap

Commercial network would use radio-sounding system.

BY ERIC HAND

Some orbiting satellites look up at the stars. Most point down towards Earth. But the satellites of the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) look sideways, across the curving horizon. There, dozens of satellites that are part of the Global Positioning System (GPS) pop in and out of view at the edge of the planet. By tracking their radio signals, COSMIC can provide atmospheric data that enhance weather forecasts and climate models.

But the fleet, launched six years ago at a cost of US$100 million, is nearing the end of its life, with one satellite of the original six already defunct. At a three-day workshop last month at the University Corporation for Atmospheric Research (UCAR) in Boulder, Colorado, researchers hailed the US–Taiwanese COSMIC as a pioneer and discussed plans for a commercial successor: a network of 24 microsatellites dubbed the Community Initiative for Cellular Earth Remote Observation (CICERO). Researchers say that the programme could help to address a gap in atmospheric data as the United States struggles to meet a 2016 launch date for the first spacecraft in its expensive Joint Polar Satellite System (JPSS). The radio-sounding technique that both COSMIC and CICERO use is a "disruptive technology", says Rick Anthes, a COSMIC scientist and former president of UCAR. "The impact is huge — especially the impact for the cost."

GPS radio signals, picked up by Earth-bound receivers in everything from mobile phones to missiles, yield precise position information. But COSMIC puts them to a different use. The signals travel at a known rate, but skimming through the planet's atmosphere and back out to space bends the signals and delays them; COSMIC uses the length of the delay to measure the atmospheric density, which can provide information on changing characteristics such as temperature and moisture levels (see 'Bending for data'). It makes many hundreds of these radio-occultation measurements each day.

Radio occultation has been a tool for planetary science from the early days of the space programme. In 1965, NASA's Mariner 4 craft broadcast radio signals to Earth across the edge of Mars, yielding estimates of the pressure of the red planet's thin atmosphere. A decade later, the Voyager probes used the technique to sample the atmospheres of Jupiter, Saturn, Uranus and Neptune.

COSMIC SUCCESS

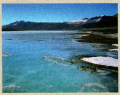
But for the technique to be useful on Earth — where basic atmospheric characteristics are already well known — a much denser network of transmitters and receivers is required. The GPS satellites offer a convenient set of transmitters, but for more than a decade the necessary receivers remained elusive. Anthes tried to gain US government support to launch COSMIC, but eventually gave up; he found a receptive audience in the then-nascent Taiwanese space programme. "It was a struggle but we did it," he says.

Since 2006, the small mission has demonstrated its many advantages. Satellites that rely on infrared sounders cannot see through clouds; microwave sounders can be confounded by intense moisture; but the radio waves seen by COSMIC pass unimpeded through even the worst storms. Furthermore, radio-occultation measurements depend only on the arrival of a signal, which can be timed using atomic clocks and easily compared between satellite systems and observing campaigns. That has made radio occultation an important calibration tool for climate scientists.

Gottfried Kirchengast, who works on radio occultation at the University of Graz in Austria, combined a decade's worth of data from COSMIC and other sources and found that

MORE ONLINE

TOP STORY



BERND WAGNER, UNIV. COLOGNE

Sealed Antarctic lake teems with life go.nature.com/cd47xr

MORE NEWS

- Benefits of breast-cancer screening disputed go.nature.com/kaqq2k
- Revolution's aftershocks still rattling Egyptian universities go.nature.com/rb1fxz
- Toxicologist Linda Birnbaum on the politics of chemical laws go.nature.com/3rm2pr

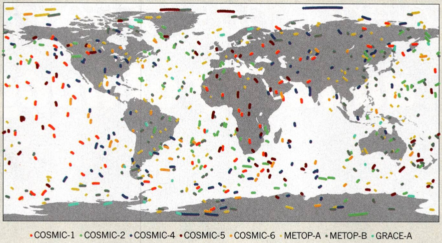
ARI FRIEDLAENDER

WATCH

Whales get in a spin for food go.nature.com/ovzrsq

A GLOBAL VIEW

Over the course of just six hours, a handful of satellites can observe radio signals as they penetrate the atmosphere at multiple locations over land and sea, providing data that blanket the globe.



SOURCE: ECMWF

atmospheric temperature was rising — evidence of anthropogenic climate change (B. C. Lackner *et al. J. Climate.* **24,** 5275–5291; 2011). “Radio occultation allowed the detection of a trend in the shortest time period ever,” he says.

Now that COSMIC has lost one satellite, says Bill Kuo, director of the programme, “We’re living on borrowed life, so to speak.” But radio-occultation data will continue to flow from the European Meteorological Operational satellite programme (MetOp), which launched a second satellite in September and is now providing about 1,400 soundings per day (see ‘A global view’). For their part, COSMIC team members hope for a successor mission, COSMIC-2, which in 2016 would launch six satellites to orbit a narrow section of the tropics, gathering data that would reduce uncertainty in measurements of hurricane intensities by 25%, and in those of hurricane tracks by 25–50%.

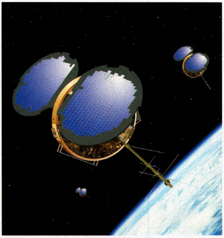
But CICERO has advantages over both of these programmes. Tom Yunck, the founder of GeoOptics in Pasadena, California, which is developing the network, says that CICERO’s fleet would observe many more occultations than COSMIC, because each member would track not only the 32 GPS satellites, but also the 24 spacecraft in the Russian Global Navigation Satellite System and the 30 satellites that will comprise Europe’s Galileo system by the end of the decade. In total, Yunck expects CICERO to generate an unprecedented 30,000 soundings per day or more.

That could help to overcome a drawback of radio-occultation data. Conventional weather satellites can offer precise horizontal resolution relative to a particular spot on Earth’s surface, whereas radio soundings provide high-resolution measurements vertically through the air column, with coarse horizontal resolution. To make up for that, CICERO will surround Earth with sensors that allow for observations along crossing sightlines, to sharpen the horizontal resolution of the measurements, and hence boost their value to forecasters.

“If you look at the impact per observation, radio occultation scores quite highly,” says Sean Healy, a scientist in the satellite division of the European Centre for Medium-Range Weather Forecasts in Reading, UK. “There are clear arguments for trying to increase the number of radio-occultation data available.”

PUBLIC–PRIVATE PARTNERSHIP

Yunck also hopes to transform the economics of weather data. Whereas most weather data are acquired with satellites paid for and launched by publicly funded agencies such as the US National Oceanic and Atmospheric Administration (NOAA), Yunck would build and launch CICERO with private funds, and then license the data to agencies in the United States and, potentially, elsewhere. Researchers would have free access to the data through the government, just as they do now. Yunck says that GeoOptics has raised $4 million of the $14 million that it needs to launch a two-satellite pilot project in mid-2014. That could expand to 24 satellites within 18 months; Yunck estimates that the total cost of building and launching the constellation would be about $150 million.



The COSMIC radio-sounding satellites are ageing but may set the stage for a commercial system.

OSC/UCAR

He hopes to prove the concept by launching CICERO before NOAA’s JPSS, a $12.9-billion programme of five weather satellites equipped with microwave and infrared sensors. The JPSS has exceeded its planned budget, and in July 2012, an independent review labelled its oversight “dysfunctional” and said that the JPSS’s delays and exorbitant costs — which the reviewers could not understand — were affecting NOAA’s credibility.

Each JPSS satellite will weigh about 2,500 kilograms. By contrast, a CICERO satellite — little more than a high-fidelity receiver — would weigh just 85 kilograms and could piggyback as a secondary payload on another launch. Yunck acknowledges that the JPSS will do things CICERO never could, such as imaging and ozone mapping. But he says that CICERO would measure temperature more precisely, and certainly more cheaply. JPSS costs “have gone through the roof, and they’re getting nowhere”, says Yunck.

In a statement, NOAA spokesman John Leslie said that, in spite of its challenges, the JPSS programme has “delivered impressive outcomes”, including the 2011 launch of a pilot craft, the Suomi National Polar-orbiting Partnership satellite. He adds that the agency “strongly supports” collecting data that will continue COSMIC’s efforts; NOAA is open to buying the data commercially, but is not convinced that commercial sources of data will become available in the short term.

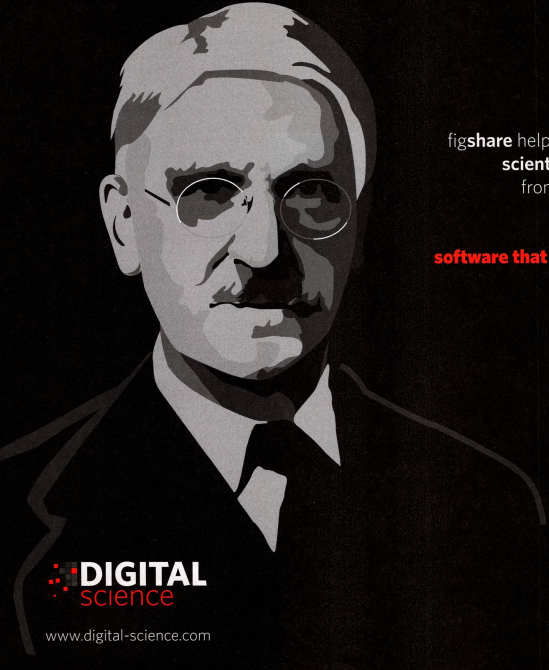
Yunck would like to change that impression. If the CICERO model succeeds, it might usher in a slew of other commercial satellite networks that could provide a range of data, including greenhouse-gas observations and gravity measurements, says Conrad Lautenbacher, chief executive of GeoOptics and a former head of NOAA. He adds, “I would like to put the government out of the business of doing routine measurements and observations that could easily be done by a commercial company.” ■

CORRECTIONS

In the News story ‘Hunt for life under Antarctic ice heats up’ (*Nature* **491,** 506–507; 2012), we wrongly stated that Martin Siegert was based at the University of Edinburgh. He is in fact based at the University of Bristol, UK. And the Editorial ‘America’s carbon compromise’ (*Nature* **491,** 301; 2012) should have talked about a $20 tax per tonne of carbon dioxide not per tonne of carbon.

"Every great advance in science has issued from a new **audacity of imagination.**"

John Dewey



fig**share** helps you **share** all of your
scientific insights and **learn**
from the insight of others.

Digital Science –
software that understands science.

DIGITAL
science

www.digital-science.com

fig**share**
credit for **all** your research

www.figshare.com

SPECIAL ISSUE

Jiangxi province, China, 2009.

REUTERS

AFTER KYOTO

In this special issue, Nature *examines the end of the 1997 Kyoto climate treaty — and the path ahead.*

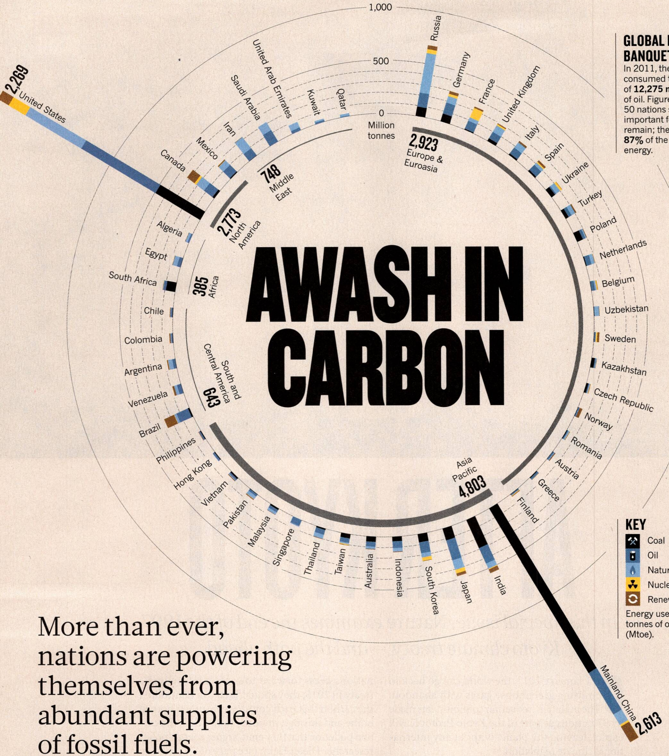
On 1 January 2013, the world can go back to emitting greenhouse gases with abandon. The pollution-reduction commitments made by 37 nations as part of the Kyoto Protocol will expire, leaving the planet without any international climate regulation.

In practice, the 1997 treaty did little to curb emissions of greenhouse gases (see page 656). Most of the parties to the treaty met their commitments easily but, because the Kyoto Protocol did not set limits for developing countries, the total emissions of greenhouse gases are rising faster than ever, thanks mainly to massive growth in coal consumption by China. A graphic view of the world's energy resources shows just how difficult it will be to wean the planet off fossil fuels (see page 654). Many nations are acknowledging the inevitable and scrambling to gird themselves against stronger and more frequent floods, droughts, heat waves and other climate threats (see page 659).

But Kyoto has provided valuable lessons. As nations press forward towards a new climate treaty in 2015, they should focus on controlling the carbon that each country consumes, both at home and through imports, rather than the carbon pollution that they emit, argues energy-policy researcher Dieter Helm (see page 663). They can also build on one legacy of the Kyoto Protocol: the carbon cap-and-trade systems and carbon taxes that have emerged in Europe, Australia, Japan, China, California and parts of Canada. Expanding these mechanisms to cover a bigger share of the world will be crucial for solving the carbon problem, says climate-policy researcher Michael Grubb (see page 666). Rather than a dead end, Kyoto could prove to be a first step towards an eventual solution. ■

AFTER KYOTO
The legacy of a climate treaty
nature.com/kyoto

AWASH IN CARBON



GLOBAL ENERGY BANQUET

In 2011, the globe consumed the equivalent of **12,275 million tonnes** of oil. Figures for the top 50 nations show how important fossil fuels remain; they supplied **87%** of the world's energy.

SOURCE: BP STATISTICAL REVIEW OF WORLD ENERGY 2012

KEY

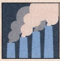
- Coal
- Oil
- Natural gas
- Nuclear
- Renewables

Energy use in million tonnes of oil equivalent (Mtoe).

More than ever, nations are powering themselves from abundant supplies of fossil fuels.

Even though countries are burning unprecedented amounts of oil and gas, the estimates of how much is left continue to grow, thanks to high prices and new technologies that have enabled companies to find and extract new resources. A decade ago, it was the tar sands of Canada and Venezuela. More recently, hydraulic-fracturing technologies have opened up oil and gas resources in the United States. Across the globe, proven oil and gas reserves are 60% higher today than they were in 1991. At current consumption rates, those reserves would last for about 60 years — and that could be extended by new discoveries and unconventional deposits. Coal reserves have not increased in size, but the supply will last for at least a century at current rates of consumption.

Renewables such as solar and wind power are growing faster than any other source of energy, but are barely making a dent in fossil-fuel consumption. The scale of the challenge will only grow as the expanding global population requires more energy. This tour of global and regional energy trends makes clear that even with aggressive action to reduce energy consumption and curb emissions, fossil fuels will be around for a very long time. ■

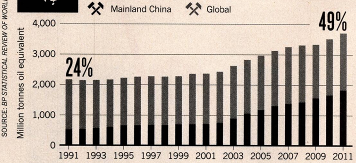


AFTER KYOTO
The legacy of a climate treaty
nature.com/kyoto

Design by
JASIEK KRZYSZTOFIAK
Reporting by
JEFF TOLLEFSON and
RICHARD MONASTERSKY

ALL THE COAL IN CHINA

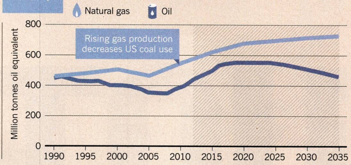
Mainland China now accounts for half of global coal consumption but at current consumption rates, it only has 33 years of domestic coal left.



SOURCE: BP STATISTICAL REVIEW OF WORLD ENERGY 2012

US BONANZA

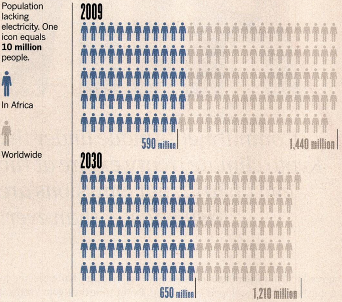
Hydraulic fracturing of tight rocks in the United States is driving a surge in oil and gas production that would continue for at least a decade in one scenario explored by the International Energy Agency.



SOURCE: IEA WORLD ENERGY OUTLOOK 2012

AFRICA UNDERPOWERED

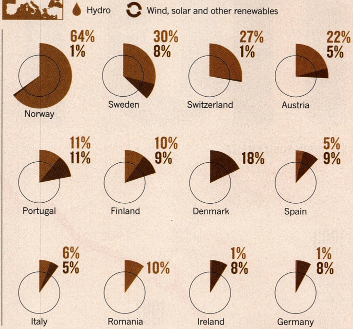
In 2009, roughly one in five people had no access to electricity, more than 40% of them in sub-Saharan Africa. By 2030, the worldwide number without electricity will fall, but more Africans will lack access.



SOURCE: IEA ENERGY POVERTY REPORT 2010

RENEWING EUROPE'S ENERGY

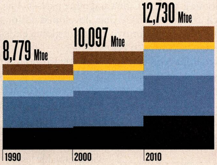
Many European nations get a sizeable fraction of their energy from renewable sources; wind and solar power are growing rapidly.



SOURCE: BP STATISTICAL REVIEW OF WORLD ENERGY 2012

PAST AND FUTURES

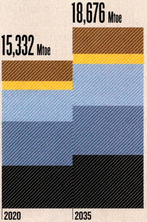
Global energy use has jumped by 45% over the past 20 years, mostly from fossil fuels. The International Energy Agency has projected demand in million tonnes of oil equivalent in three scenarios. Each scenario has very different consequences for greenhouse-gas changes between 2010 and 2035. Keeping carbon dioxide levels below 450 parts per million (p.p.m.) would give the world even odds of limiting global warming to 2 °C.



SOURCE: IEA WORLD ENERGY OUTLOOK 2012

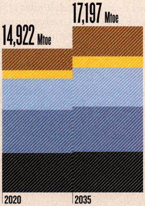
CURRENT POLICIES CONTINUE

Emissions rise **46%**



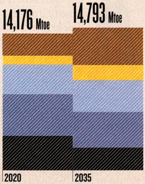
NATIONS CARRY OUT ANNOUNCED ENERGY POLICIES

Emissions rise **23%**



NATIONS AIM TO KEEP CO_2 UNDER 450 P.P.M.

Emissions fall **27%**

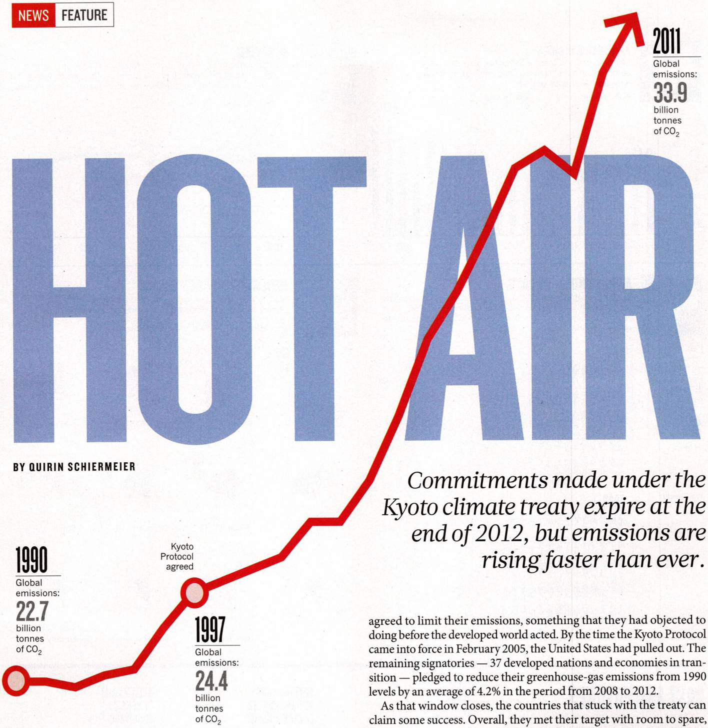


NEWS FEATURE

HOT AIR

BY QUIRIN SCHIERMEIER

Commitments made under the Kyoto climate treaty expire at the end of 2012, but emissions are rising faster than ever.



SOURCE: NETHERLANDS ENVIRON. ASSESSMENT AGENCY/EC JOINT RES. CENTRE

After 8 days of fractious negotiating, delegates at the 1997 climate conference in Kyoto, Japan, were running out of time to deliver a treaty aimed at slowing global warming. The leader of the talks, Michael Zammit Cutajar of Malta, took the unusual step of invoking Zen Buddhism, telling everyone that they must break through mental barriers to achieve enlightenment. Two days later, after a marathon all-night session, the negotiators finally hammered out the climate agreement known as the Kyoto Protocol. It was the first — and so far, only — pact to commit rich countries to reducing emissions of carbon dioxide and other greenhouse gases.

But even before the ink was dry on the agreement, it was clear that the protocol faced a rocky future. Although the United States had signed the treaty, President Bill Clinton signalled that the world's largest economy would not ratify the pact unless China and other developing nations agreed to limit their emissions, something that they had objected to doing before the developed world acted. By the time the Kyoto Protocol came into force in February 2005, the United States had pulled out. The remaining signatories — 37 developed nations and economies in transition — pledged to reduce their greenhouse-gas emissions from 1990 levels by an average of 4.2% in the period from 2008 to 2012.

As that window closes, the countries that stuck with the treaty can claim some success. Overall, they met their target with room to spare, cutting their collective emissions by around 16%. But most of those cuts came with little or no effort, because of the collapse of greenhouse-gas producing industries in eastern Europe and, more recently, the global economic crisis.

Furthermore, the cuts by industrialized nations have done little to combat the global problem. Worldwide emissions have surged by 50% since 1990, driven by economic growth in China and other parts of Asia, South America and Africa. In the 1990 base year, developed nations including the United States accounted for two-thirds of global emissions. Now, their contribution has dropped below 50%.

"Kyoto had a very limited impact on climate," says Atte Korhola, an environmental-policy researcher at the University of Helsinki. "It was too narrow in ambition, its tools were too massively bureaucratic and it offered too many loopholes."

But the treaty has taught policy-makers some valuable lessons

AFTER KYOTO
The legacy of a climate treaty
nature.com/kyoto

and possibly laid the groundwork for more ambitious efforts. "Kyoto was a grand policy experiment with important lessons we ought to take forward. It had its flaws — no wonder, you rarely get policies right the first time — but the overall architecture is still useful," says Roger Pielke Jr, who studies energy and innovation policy at the University of Colorado Boulder.

DIFFICULT LEGACY

The seeds of Kyoto's problems were planted long before the treaty took shape. Many go back to June 1992, when negotiators at the Earth Summit in Rio de Janeiro, Brazil, were hammering out the United Nations Framework Convention on Climate Change (UNFCCC), the umbrella treaty that would encompass the Kyoto Protocol. Negotiators in Rio were still crafting the document just hours before heads of state arrived to sign it. Pressed by time and mounting expectations, the delegates borrowed heavily from past treaties, including a US–Soviet nuclear-arms agreement and the 1989 Montreal Protocol designed to protect the ozone layer, says Gwyn Prins, who studies environmental politics at the London School of Economics and acted as an adviser for the British negotiating team in 1992.

"Take out nuclear warheads, put in CO_2 — the basic idea was as easy as that," says Prins. "But it turned out that climate change is a much more wicked beast — scientifically and economically — than ozone chemistry or nuclear-arms control."

> "KYOTO WAS A GRAND EXPERIMENT WITH IMPORTANT LESSONS WE OUGHT TO TAKE FORWARD."

A meeting in Berlin in 1995 created another major problem, when parties to the UNFCCC decided to divide the world into two categories for the future treaty. There would be a set of rich countries with ambitious climate responsibilities and a set of less-developed economies — including China — with no responsibilities.

That decision, part of an agreement known as the Berlin Mandate, did not sit well with US politicians. In the summer of 1997, Robert Byrd, a Democratic senator from West Virginia and one of the senior politicians of his day, declared: "It is the Berlin mandate — and the fact that it lets the developing world off the hook scot-free — that will seriously harm the global environment in future years."

His colleagues agreed. The US Senate voted 95 to 0 in favour of a proposal demanding that developing nations participate in emissions commitments. Because Kyoto included no such commitments, the United States — the world's largest greenhouse-gas emitter at the time — would not ratify it.

The industrialized countries that remained with the treaty were each bound by individualized commitments, based on the state of their economy and energy mix at the time (see 'Uneven progress'). The developed nations of Germany and Denmark agreed to cut their emissions by 21% relative to 1990 levels, whereas Portugal, with its less-developed economy, was allowed to increase its emissions by 27%.

Kyoto covered four main greenhouse gases — CO_2, methane, nitrous oxide and sulphur hexafluoride — and two further groups of gases, hydrofluorocarbons and perfluorocarbons. But it did not include another warming force: black soot particles from the incomplete combustion of wood and fossil fuels.

Countries could meet their commitments by cutting their own emissions or by buying emission allotments from other nations that had exceeded their required reductions. Rich countries could also get credit by investing in low-carbon technologies in developing countries.

For most central and eastern European nations, the job was easy: industrial emissions were high in the base year but had plummeted even by the time the treaty was signed. By 2010, Russia's CO_2 emissions were 34% lower than in the base year (excluding cuts attributable to land-use changes) and Ukraine's had fallen by 59%. The United Kingdom also easily met its 12.5% reduction target, thanks to the closure of many coal mines and a corresponding drop in consumption.

More recently, the economic downturn has helped to reduce emissions. Economists estimate that between 2007 and 2008, decreased energy use caused a 2% drop in the emissions of the Kyoto Protocol countries; and that trend has continued as economies have sputtered.

But the reductions made under the treaty were dwarfed by the rise in emissions not covered by the accord, especially in Asia. Since 2000, CO_2 emissions in China have nearly tripled to almost 10 billion tonnes, and those in India have doubled to around 2 billion tonnes.

The rise in Asian emissions is partly a result of the migration of heavy industry from developed nations to developing countries, which make products that then get shipped back to wealthy nations. Between 1990 and 2010, the emissions embodied in such products grew by an average of 10% per year — to an annual total of 1.4 billion tonnes — surpassing the total emissions reductions achieved under Kyoto, says Glen Peters, a climate-policy researcher at the Center for International Climate and Environmental Research — Oslo. The gains made by the treaty were therefore deceptive, says David Victor, an energy-policy researcher at the University of California, San Diego. The treaty, he adds, was based on "dubious economic assumptions and flawed accounting systems".

FAULTY REASONING

One of those dubious assumptions was that fossil fuels would soon grow scarcer and prices would spiral upwards, helping to push countries towards alternative energy sources. But the globe is currently going through a massive coal renaissance, driven by abundant supplies that have grown much cheaper relative to other fuels in much of the world: the share of energy derived from coal has increased in the past ten years in both developing and developed countries. There has even been a shift towards coal in some parts of Europe, despite the mandatory cap-and-trade system to limit emissions. As a result, global energy production has grown more carbon-intensive in the past decade.

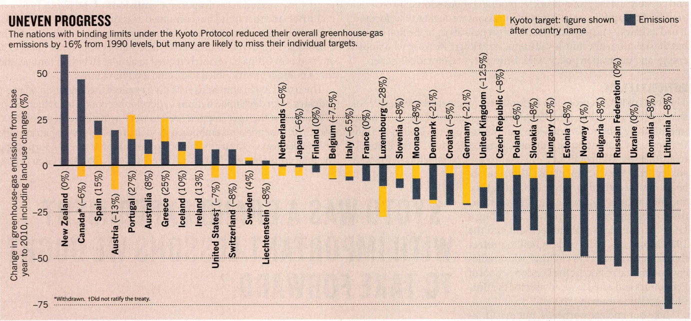
"The fathers of the UNFCCC and Kyoto Protocol quite severely underestimated the amount of hydrocarbons buried in the ground," says Ottmar Edenhofer, chief economist at the Potsdam Institute of Climate Impact Research in Germany and a lead scientist with the Intergovernmental Panel on Climate Change.

These trends in energy use have made it nearly impossible for countries to limit global warming to less than 2 °C above preindustrial levels, the value chosen by the EU as a threshold likely to prevent dangerous climate change. Calculations suggest[1] that emissions of CO_2 must stay below 1,000 billion tonnes between 2000 and 2050 to give the world a 75% chance of containing the temperature rise to 2 °C.

But emissions from fossil-fuel burning and deforestation since 2000 have already pumped more than 450 billion tonnes of CO_2 into the atmosphere. If the current trend continues, the 1,000-billion-tonne margin will be surpassed in a little more than a decade.

Despite its shortcomings, Kyoto has not been an utter failure, says Robert Stavins, an environmental economist at Harvard University in Cambridge, Massachusetts. Rather than judging the agreement on the emissions reductions it has achieved, he says, people should consider whether it has put the world on the right path.

"Nobody with a right mind could have expected that a climate regime that treats China like sub-Saharan Africa and that excludes 50 developing countries with a higher per-capita income than Romania could be anything other than a cautious first step," he says. "What we need to create is a workable successor with binding national emission targets that all governments can be realistically expected to adopt."



UNEVEN PROGRESS

The nations with binding limits under the Kyoto Protocol reduced their overall greenhouse-gas emissions by 16% from 1990 levels, but many are likely to miss their individual targets.

*Withdrawn. †Did not ratify the treaty.

SOURCE: NETHERLANDS ENVIRON. ASSESSMENT AGENCY/EC JOINT RES. CENTRE

Kyoto will leave a valuable legacy, says Yvo de Boer, former chief executive secretary of the UNFCCC and adviser for global auditing firm KPMG. The methodologies developed for reporting and verifying national greenhouse-gas emissions and land-use changes will be important components of any future climate treaty, he says.

The protocol also gave birth to a method for trading carbon emissions among countries that face limits. Pioneered by the EU's Emissions Trading Scheme, which launched in 2005, this carbon market could one day become a globally linked CO_2 cap-and-trade system, says de Boer.

An additional element of the Kyoto agreement — the Clean Development Mechanism (CDM) — established a way for rich countries to get credits towards their targets by making cost-effective emissions cuts in poor countries. Critics have charged that the CDM is plagued by cumbersome bureaucracy and that some Western-funded clean-technology projects in developing countries would probably have been built without it. Nevertheless, a total of 5,000 CDM projects have attracted investments worth almost US$100 billion. The projects have ranged from providing rural Chinese villagers with solar cookers to supporting a 100-megawatt wind farm in Mexico.

"Without Kyoto we wouldn't have achieved anything at all" in that area, says Victor. He would like to see a successor treaty constructed more like trade accords, which are tailored using realistic assumptions about commitments and rely on mutual action. "What one country is willing to pay to control emissions depends a lot on what its economic competitors will pay as well," he says. "More flexible treaties could help countries craft deals that are truly interdependent — where the efforts of one country get multiplied because they lead others to do more."

FOLLOW THE MONEY

Many other policy experts agree that the next climate treaty must take a more pragmatic approach than the UNFCCC and the Kyoto Protocol, which failed to win over the biggest polluters in part because it relied on a mix of ethical and environmental rationales rather than economic ones. "Making energy more expensive is a political liability everywhere," says Pielke. "When emission reductions run up against economic growth, economic growth will inevitably win out. There is no magical solution, so you better set yourself tangible goals that aren't doomed to clash with the iron laws of politics."

Emissions targets for all countries should be allocated in a way that acknowledges the political and economic costs of complying with a climate agreement, argued Valentina Bosetti, a climate-impact modeller at the Eni Enrico Mattei Foundation in Milan, Italy, and Jeffrey Frankel, an economist at Harvard, in a discussion paper last year[2]. China, for example, would be asked to accept only targets that it could meet without sacrificing its developmental aspirations; the United States would be assigned more stringent goals. But with time, all nations' emissions targets would be adjusted progressively according to a common economic formula.

Attaching a price to carbon, through cap-and-trade mechanisms or a direct carbon tax, would help by stimulating technological advances that reduce emissions. The challenge, says Pielke, is to get the price right and make sure that the revenue will go towards investments in technology.

A moderate carbon tax — applied when fossil fuels are removed from the ground — might work best to stimulate innovation in technologies that will eventually make alternative energy sources cheaper than fossil fuels, he says. But the approach has to be global.

In a policy paper[3] published in 2010, Pielke, Prins and 12 others called for a more pragmatic, diversified and less bureaucratic approach than Kyoto, which would wean the global economy off carbon as a by-product of reducing poverty and expanding energy access to the poor.

The group takes the focus off CO_2, which has a long lifetime in the atmosphere, and instead emphasizes cuts in black carbon and methane emissions, which don't last as long. This, say the paper's authors, would slow global warming more quickly and would provide time for a transition to a low-carbon economy. They also suggest that negotiations for the next emissions treaty avoid topics such as deforestation, land use, air quality and adaptation, which would greatly complicate its architecture.

That agreement will take shape slowly over the next few years. In Copenhagen in 2009, nations failed to produce a follow-on treaty to the Kyoto Protocol. However, in Durban, South Africa, last year, countries including China and the United States agreed to negotiate a new climate treaty by 2015. If the past is any indication, the final details of that pact will not emerge until the sleep-deprived delegates have reached the deadline of the final negotiating session.

Will the world find a solution to this so far intractable problem? "I'm confident it will," says de Boer, who presided over the unsuccessful negotiations in Copenhagen. "But I'm not convinced that it will come on time." ■

Quirin Schiermeier *is a reporter for* Nature *in Munich, Germany.*

1. Meinshausen, M. *et al. Nature* **458**, 1158–1162 (2009).
2. Bosetti, V. & Frankel, J. *Sustainable Cooperation in Global Climate Policy: Specific Formulas and Emission Targets to Build on Copenhagen and Cancun* Discussion paper 2011-46 (Harvard Project on Climate Agreements, 2011).
3. Prins, G. *et al. The Hartwell Paper: A New Direction for Climate Policy After the Crash of 2009* (LSE, 2010).

Bangladeshis use the ubiquitous hyacinth weed to build floating, flood-proof crop gardens.

JONAS BENDIKSEN/MAGNUM PHOTOS

NO GOING BACK

With nations doing little to slow climate change, many people are ramping up plans to adapt to the inevitable.

BY OLIVE HEFFERNAN

When Superstorm Sandy hit the US coast last month, it blew millions of New Yorkers back into the nineteenth century. The southern part of Manhattan went black after floodwaters shorted out electrical systems. With the subway system disabled, many residents resorted to traversing the island by foot, and water supplies in some areas became contaminated with bacteria and pollutants.

The largest Atlantic hurricane on record, Sandy wreaked US$50 billion in economic losses along the US northeast coast, providing a costly reminder of how ill-prepared even the richest nations are for weather extremes. Some recent weather disasters have now been attributed, at least in part, to human activity, including the 2003 European heatwave[1] and the floods in England in 2000 (ref. 2). According to the Intergovernmental Panel on Climate Change (IPCC), storms, floods and droughts will strike more frequently and with greater strength as the climate warms[3]. And if nations are struggling to cope now, how will they manage in a warmer, harsher future?

Just a decade ago, 'adaptation' was something of a dirty word in the climate arena — an insinuation that nations could continue with business as usual and deal with the mess later. But greenhouse-gas emissions are increasing at an unprecedented rate and countries have failed to negotiate a successor to the Kyoto Protocol climate treaty. That stark reality has forced climate researchers and policy-makers to explore ways to weather some of the inevitable changes.

"As progress to reduce emissions has slowed in most countries, there has been a turn towards adaptation," says Jon Barnett, a political geographer at the University of Melbourne in Australia.

Adaptation has tended to focus on hard defences, such as shoring up sea walls and building dams. But as awareness of adaptation has grown, so too has the concept. "Adaptation means different things to different people, and is extremely location specific," says Neil Adger, an environmental and economic geographer at the University of Exeter, UK. Although residents in Bangladesh can raise their houses on stilts to survive floods, some settlements in Alaska and the Maldives must move in the face of rising sea levels.

Increasingly, it is local people who are deciding how to make their communities more resilient — and that is increasing the chances

AFTER KYOTO
The legacy of a climate treaty
nature.com/kyoto

of success. “A solely top-down approach to adaptation — focusing on heavy investment in engineering and infrastructure — will not work as it is expensive and impractical,” says Robert Lempert, who researches decision-making at the RAND corporation, a think tank in Santa Monica, California.

FEARSOME FLOODS

With its low-lying deltas, Bangladesh has always been threatened by storms that blow in from the Bay of Bengal. But the cyclone that hit the southeastern edge of the nation on 29 April 1991 caused massive flooding and killed some 138,000 people, mainly children and older women. Twenty-one years earlier, a cyclone had claimed up to a half million lives.

Rising sea levels are increasing the risks to Bengalis, both from cyclones and from the spread of saline groundwater, which ruins aquifers and kills off crops. Projections by the IPCC indicate that a sea-level rise of 1 metre — expected sometime within a century — would inundate up to 20% of the country’s land area and displace 14% of the total population[4].

In response, the country is busy building and strengthening its 13,000-kilometre network of embankments, planting salt-resilient crops and storing fresh water. One project, supported by the United Nations Development Programme, taught 18,000 households in coastal communities to plant mangroves and fruit trees and to harvest rainwater by digging ditches. The project aims to provide fresh water and income as well as to protect against flooding and erosion. In 2005, the country became one of the first to complete a national adaptation programme of action, and it later established a climate-change trust fund for financing local, small-scale projects in adaptation and mitigation.

Although Bangladesh’s per capita income ranks the nation among the world’s poorest, the country has mounted such a strong response to climate change that it has become something of a model for other nations, says Saleemul Huq, a Bangladeshi scientist and senior fellow in the Climate Change Group at the International Institute for Environment and Development in London. “In the past few years, the level of public awareness about climate change in Bangladesh has skyrocketed and it is now probably higher there than in any other country,” Huq says. But here and in other developing countries, it is hard to separate adaptation efforts from development that would have happened anyway. This confusion permeates discussions on financing adaptation efforts; of the $125 million that Bangladesh has received in climate funds from overseas, it remains unclear how much is in addition to development aid.

“A SOLELY TOP-DOWN APPROACH TO ADAPTATION — FOCUSING ON HEAVY INVESTMENT IN ENGINEERING AND INFRASTRUCTURE — WILL NOT WORK.”

And like almost everywhere, Bangladesh is having to play catch-up with the climate. “There is an adaptation deficit out there right now to current climate variability,” says Kristie Ebi, an expert on climate and health impacts at Stanford University in California.

Still, many experts say that Bangladesh has made significant progress. “What we can say is that there is collective action in the most climate-vulnerable country in the world,” says Huq. “The issue has galvanized people even across the political spectrum.”

Whereas Bangladesh has focused on involving citizens in many small-scale projects, the city of Melbourne has sought to head off problems through a massive engineering venture. Over the past decade, southern Australia has been hit by the worst drought in a century. After water restrictions implemented in 2007 angered thousands of farmers, the state of Victoria announced it would invest Aus$3.5 billion (US$2.9 billion) in a new desalination plant at the site of Wonthaggi and commissioned a pipeline to bring water to the region from a river in the north. Overdue and over budget, the desalination project will take decades to pay off and is eating up the region’s adaptation resources, argue critics.

It is a clear case of maladaptation and will increase overall vulnerability to climate change, says Barnett, who has studied the project[5]. By investing so heavily in the desalination plant, the Victorian government has effectively shut off the possibility of funding other adaptation options, such as harvesting rainfall and recycling domestic waste water from showers and dishwashing, which would be cheaper and more effective, he says.

But John Thwaites, the water minister of Victoria until mid-2007, says that Melbourne has taken a multi-pronged approach that includes conservation and harvesting storm water. The government originally preferred measures other than desalination, but decided to build the plant in response to the unprecedented water shortage.

The United Kingdom has taken a different approach to planning for water problems. In autumn 2000, rainfall was at its highest since records began in 1766, causing devastating floods that left the village of Hambledon under water for six weeks at a cost of more than £1 billion (US$1.6 billion). Early this year, a long drought led to water restrictions in parts of Britain, but the spring and summer that followed were the wettest in a century. Climate projections indicate that the country will face more frequent droughts and floods, but the models do not agree on the extent and timing of the changes.

Given the uncertain predictions, UK water companies are adopting a more flexible approach for making decisions about building reservoirs, extracting groundwater and other water-related plans. Rather than developing one strategy and sticking with it, they plan incrementally and frequently re-evaluate on the basis of new information, says Nigel Arnell, director of the climate research-focused Walker Institute at the University of Reading, UK.

Communities living on tiny islands, however, don’t have the luxury of considering many different options and reevaluating plans. For people in the Maldives, Kiribati and Carteret, there is simply nowhere to retreat when rising sea levels infiltrate their drinking-water supplies and flood their homes, so they will have to flee.

Small island nations can learn valuable lessons from the Alaskan village of Newtok, which is already in the process of moving. Located on the Ninglick River next to the Bering Sea, Newtok is below sea level and losing ground at a rate of roughly 22 metres per year to erosion. The villagers selected a site for their new home on Nelson Island, which is 15 kilometres away. Turning a hardship into an opportunity, the residents are learning building skills to construct sustainable homes in their new village, and that will give them more job options in the future, says Robin Bronen, head of the Alaska Immigration Justice Project in Anchorage, which works with communities being relocated as a result of climate change in Alaska.

The experience in Newtok serves as a model for how ‘climigration’ should work in practice, says Bronen. “That’s because of the process — the community has made all of the decisions.”

Heat is often considered less dangerous than floods, but some of the most serious consequences of climate change will arise from hot weather. During the 2003 heatwave in Europe, temperatures reached their highest since 1500, topping 40 °C for seven days in some parts of France. Almost 15,000 people died in that country alone. One climate modelling study of the Mediterranean region found that by the end of the current century, the frequency of heatwaves will increase from one every 3–5 summers to 2–3 heatwaves each summer; and they will last 2–5 times longer[3].

In the aftermath of the 2003 disaster, France established a heatwave warning system, one of 12 now in operation throughout Europe[6]. In France, alerts are triggered when the five-day weather forecast predicts that temperatures will exceed thresholds for three days. In addition, scientists analyse mortality data, hospital patient loads and drinking-water

The costly Wonthaggi Desalination Plant in Australia has been criticized for consuming adaptation resources.

CARLA GOTTGENS/BLOOMBERG VIA GETTY

supplies. The system can issue public warnings, mobilize personnel to visit vulnerable populations and call hospital and nursing-home staff back from their holidays. In 2006, a heatwave similar to the 2003 event put the system to the test; although it did reduce mortality and morbidity, thousands still died, in part because a significant fraction did not receive the warning and many did not heed the advice.

Ebi says that cities can do much better. "Heat-related deaths are completely preventable," she says, if people are warned and told how to protect themselves. Yet the most vulnerable — the elderly, ill and poor, for example — often don't see themselves as being at risk, says Graham Bickler, regional director at the Health Protection Agency in Brighton, UK. So one challenge is identifying and communicating to those groups. During a heatwave in Chicago, Illinois, in 1995, Ebi says, many of those who died were adults over 65 who lived in poor areas, where it was common practice to board up windows to protect against break-ins. The fans they used to try to keep cool had the reverse effect and turned their homes into convection ovens. They could have reduced their risk by consuming cold liquids and foods, taking cool showers and going to air-conditioned public spaces, says Ebi.

EARLY WARNING

Like Bangladesh, the nations of sub-Saharan Africa are particularly vulnerable to climate change, but they have received relatively little adaptation funding from international donors. Mozambique stands out as one of the most threatened, with 2,700 kilometres of coastline and more than half its 24 million inhabitants living in poverty.

Between 1965 and 1998 the country experienced 12 major floods, 9 major droughts, and 4 major cyclone events. Located at the end of river systems that stretch more than 1,000 kilometres into other countries, Mozambique can be struck floods without warning. "Mozambique is a country where every year there is a weather-related problem," says Filipe Lúcio, a Mozambican meteorologist who now heads the Global Framework for Climate Services Office at the World Meteorological Organization in Geneva, Switzerland.

In 2000, Mozambique was hit by a flood worse than any in its history; it left 700 dead and caused damages totalling nearly US$300 million, a significant fraction of its budget at the time. The event "wasn't at all anticipated", says Lúcio. Warnings of above-average rainfall came too late and failed to convey the magnitude of the coming flood. Since then, the country has switched to a colour-coded system that indicates the lead time for a predicted event and has recruited locals to record and monitor precipitation and water levels — information they can use to raise alarms.

The system has succeeded in reducing risks. In 2007, a flood of similar magnitude to the 2000 event killed just 29 people. Looking to the future, the government has commissioned a study on national climate impacts and is working on a long-term strategic adaptation plan. But efforts to implement these plans will probably be hindered by lack of funding and sluggish bureaucracy[7]. The state has a budget of just $5 billion, and half of Mozambique's funds come from overseas aid, so resources are scarce even for crucial areas such as education and health. As of November 2011, Mozambique had received $30 million in overseas climate finance and a further $86 million has been pledged by the Pilot Program on Climate Resilience for a range of issues from upgrading roads to improving its meteorological service[8]. But experts say that this is far from enough.

It is a picture repeated across the globe. In 2011, developing nations received only about $960 million in money dedicated for adaptation-related activities, but a 2007 report by the United Nations Development Programme estimated that developing countries would need $86 billion a year by 2015 in funding to adapt to climate change[9].

The issue is pushing the world's rich and poor farther apart, says Desmund Tutu, the former archbishop of Cape Town in South Africa. In the UN report, he warned that "we are drifting into a world of 'adaptation apartheid'".

For both wealthy and poor nations, the challenge is to convince people to act before it is too late. "Adaptation is largely a matter of changing social processes so that fewer people are at risk," says Barnett. "This of course won't be easy — but it does mean the solutions are determined by people, not by nature". ■

Olive Heffernan *is a freelance writer and the former editor of* Nature Climate Change.

1. Stott, P. A., Stone, D. A. & Allen, M. R. *Nature* **432,** 610–614 (2004).
2. Pall, P. *et al. Nature* **470,** 382–385 (2011).
3. Fisher, E. M. & Schär, C. *Nature Geosci.* **3,** 398–403 (2010).
4. Intergovernmental Panel on Climate Change *Climate Change 2007: Impacts, Adaptation and Vulnerability* Ch. 10 (Cambridge Univ. Press, 2007).
5. Barnett, J. & O'Neill, S. *Glob. Environ. Change* **20,** 211–213 (2010).
6. Lowe, D., Ebi. K. L. & Forsberg, B. *Int. J. Environ. Res. Public Health* **8,** 4623–4648 (2011).
7. Blythe, J. *Hits and misses in Mozambique's climate change action plans* Africa Portal Backgrounder no. 7 (2012).
8. Nakhooda, S., Caravani, A. & Bird, N. *Climate Finance Regional Briefing: Sub-Saharan Africa* (Overseas Dev. Inst., 2011).
9. United Nations Development Programme *Human Development Report 2007/2008* (UNDP, 2007).

eppendorf

Winner 2012
Elizabeth Murchison,
United Kingdom

APPLICATION DEADLINE
15 JANUARY 2013

Call for Entries!

Eppendorf Award 2013 for Young European Investigators
Since 1995 the Eppendorf Young Investigator Award is granted annually to European researchers not older than 35 years. It acknowledges outstanding contributions to biomedical research in Europe based on methods of molecular biology, including novel analytical concepts. The winner is selected by an independent expert committee chaired by Reinhard Jahn (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany).

Selection criteria
1. Ability to work independently
2. Creativity
3. Successful practical application of research approach
4. Significance of research results for the field of medicine
5. Achievement of candidate in relation to his/her age

You could be next to receive
> a prize money of 15,000 Euro
> an invitation to the prize ceremony at the EMBL Advanced Training Centre in Heidelberg on 6 June 2013
> an invitation to visit Eppendorf AG in Hamburg
> coverage of your work by Nature in print and in a podcast

Apply online at www.eppendorf.com/award/application

The Eppendorf Award is presented in partnership with
nature

Eppendorf® is a registered trademark of Eppendorf AG. All rights reserved, incl. graphics and images. Copyright ©2012 by Eppendorf AG.

COMMENT

KYOTO Smaller carbon-pricing schemes could keep emissions in check **p.666**

ENGINEERING A life of the man who made skyscrapers withstand wind **p.668**

CULTURE Osseous overload links two London exhibitions on death **p.671**

CONSERVATION Europe's ash dieback could leave many lichens homeless **p.672**

JIANAN YU/REUTERS

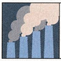
Around 80% of China's electricity generation is coal-fired.

The Kyoto approach has failed

Abandon coal, price carbon consumption and look to new technologies for a lasting solution to global emissions, argues **Dieter Helm**.

The Kyoto Protocol, agreed in 1997, is the centrepiece of global efforts to address climate change by reducing greenhouse-gas emissions. Its first commitment period expires this year, but despite the political capital invested in it, numerous subsequent Conference of the Parties (COP) meetings and considerable economic costs, it has had no noticeable impact on global carbon emissions. These remain on an upward curve, increasing from almost 2 parts per million (p.p.m.) a year in the early 1990s to almost 3 p.p.m. now, and heading towards the critical threshold of 400 p.p.m..

It will only get worse. At the Durban COP in December 2011, all that could be agreed was that the participant countries would try to agree by 2015 what they might do after 2020. At current growth rates, by 2020 the economies of China and India will be twice their present size, requiring the addition of 400–600 gigawatts of coal-fired generating capacity to their electricity systems[1].



AFTER KYOTO
The legacy of a climate treaty
nature.com/kyoto

The reasons for the Kyoto Protocol's ineffectiveness are in its architecture. It is based on carbon production, not carbon consumption. It has a mainly European focus. It does nothing to address ▶

▶ the immediate problem of global coal burning. It is wide open to free-riding, allowing nations to avoid cutting emissions while others do so, and it has few enforcement mechanisms. These are deep flaws that render the protocol incapable of slowing emissions, let alone reversing them. Fortunately, other, better, bottom-up approaches hold hope for progress.

CARBON FOOTPRINT

The idea at the heart of the Kyoto Protocol is that the developed countries accept caps on carbon production from power stations, industrial installations and the like within their borders. Developing countries take measures but need not apply caps. On aggregate, carbon emissions should have been reduced by about 5% below 1990 levels by the end of 2012.

The main problem with the Kyoto approach is that it does not address the carbon footprint — carbon consumption. A country's (and an individual's) carbon footprint is best measured by looking at the carbon embedded in the goods and services that each consumes. Global warming takes no account of national boundaries. If a US consumer buys a car, it matters little whether the steel within it is made in the United States or China.

The difference between carbon production and carbon consumption is not trivial. Take the United Kingdom: from 1990 to 2005, its carbon production fell by around 15%. But carbon consumption went up by around 19% once the carbon embedded in imports is taken into account[2].

From a Kyoto perspective, this is a triumph; for climate change, it is a disaster. It explains how emissions can apparently fall in Europe but go up globally as rapidly developing countries, such as China and India, export energy-intensive goods to Europe and the United States, which together make up around 50% of the world's gross domestic product.

It is not surprising that Europe has led the way on Kyoto. Carbon-production targets have been comparatively easy to meet, and they make Europe look good. But the real reasons for the fall in carbon production do not give much cause for celebration.

The collapse of the Soviet Union began at the end of the 1980s — nicely timed for the use of 1990 as Kyoto's reference baseline year (see 'Carbon climb'). Before this, Eastern Europe was notorious for inefficient, energy-intensive industrial production, much of which — from a Kyoto perspective — conveniently stopped after the fall of the Berlin Wall. Once the United States had opted out of the Kyoto Protocol, the agreement needed Russia to come on board so that it could come into force; it was a condition that the protocol should be ratified by at least 55 countries, covering 55% of global emissions in 1990. Russia brought lots more 'hot air' — emissions reductions that were inevitable.

Better still from the perspective of Kyoto compliance, western Europe was de-industrializing too, switching away from energy-intensive production activities towards service industries, in part because Chinese exports were outcompeting industries in Europe.

Most of this would have happened anyway. Europe's green policies have made little difference, and the economic crisis has made reducing its carbon production even easier. The 2008 climate-change package from the European Union (EU) focuses on the short term[3]. By 2020, it aims to reduce EU carbon emissions by 20%, to increase energy generation from renewables by 20% and to boost energy efficiency by 20%. Aside from the economic illiteracy of assuming that everything adds up to the magic number of 20, the effect has been to focus almost all resources on a small number of current renewable-energy technologies — wind, rooftop solar and biomass.

> ***"The real villain of growing global emissions has been ignored: coal."***

These measures were intended to reinforce the EU Emissions Trading Scheme (EU ETS)[4], which has produced a short-term, volatile and low price for carbon when what is required is a medium- to long-term, stable but rising carbon price. The net effect of all these EU measures (especially the renewables) has been to drive up energy prices and reduce European competitiveness, while making almost no contribution towards mitigating global emissions.

THE COAL PROBLEM

The real villain of growing global emissions has been ignored: coal. Since the mid-1990s, coal has risen from supplying around 25% of the world's primary energy to almost 30% now, in a context of a rapidly growing underlying energy demand[5].

Much of this coal burning has been in China, which switched from being an exporter to an importer of coal in the 1990s, and now accounts for a staggering 50% of world coal trade. Its share of global coal production is almost four times that of Saudi Arabia's production of oil[6]. Around 80% of China's electricity generation is coal-fired. China and India together add around three coal-fired power stations a week to their generation portfolios.

China plans to improve its energy efficiency, reducing its energy intensity under its 12th Five Year Plan by 16% between now and 2015. It intends to develop gas and renewables. But the arithmetic of an economy that doubles in size every decade puts these changes into perspective: a slightly smaller proportion of coal-fired electricity generation in an economy twice the size by 2020 becomes lost in the noise. The world faces a further huge increase in coal burning between now and 2020, and the result will be ever-rising emissions. The Kyoto Protocol has almost nothing to say about this.

The story on coal elsewhere is mixed. In the United States — which remains outside the Kyoto Protocol — carbon emissions have been falling sharply[7], and faster than in crisis-ridden Europe over the past five years. The United States is shifting from coal to natural gas for electricity generation and industry as the full impact of the shale-gas revolution is played out. Natural gas produces fewer pollutants and half the carbon emissions of coal.

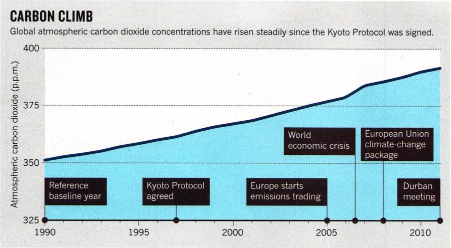
With the price of natural gas in the United States currently around one-quarter of that in Europe, and even lower than that in China, economics is driving a shift towards natural gas without the need for any drastic energy or climate-change policies. So great is the competitive advantage bestowed by shale gas that energy-intensive industries are beginning to migrate from China back to the United States. Ironically, these repatriated industries will push up US carbon production while reducing China's. The net effect is good from a global climate-change perspective — less transport to the United States and more-efficient electricity generation — but it would make the country look bad by Kyoto's emissions-cap approach.

In Europe, the irony is deeper still — many countries are switching from nuclear and gas to coal. Germany stands out. It has prematurely closed some of its existing nuclear-power stations, and is fast-tracking the closure of the rest within the next decade. It is getting out of low-carbon generation in a big way. Germany is increasing the burn in its existing coal power stations, and building new ones that burn lignite, one of the dirtiest forms of coal. The use of natural gas is being squeezed out by low coal prices across Europe, and because the EU ETS carbon price is so low as to be negligible.

FREE-RIDING

Advocates of the Kyoto approach argue that these problems are temporary. Over time, other countries will join, eventually resulting in a complete set of carbon-production caps. Then the distinction between carbon production and carbon consumption will not matter. Putting aside the facts — that nothing much is going to happen until at least 2020, and that we don't have the luxury of waiting as emissions pile up — why should we have any confidence in the gradual evolution of Kyoto?

The core architecture of the protocol relies



CARBON CLIMB

Global atmospheric carbon dioxide concentrations have risen steadily since the Kyoto Protocol was signed.

SOURCE: US NATL OCEANIC AND ATMOSPHERIC ADMINISTRATION

on countries adopting emissions caps and complying with them in the face of major free-rider incentives, differential impacts and no serious enforcement mechanisms. Kyoto is similar to the classic prisoners' dilemma in game theory. Each country may be better off if all nations cooperate, but each may be tempted to free-ride on the costs and efforts of others.

For instance, emissions from some countries are capped, but those from others are not. The result is that the implicit (or explicit) price of carbon in countries with a cap creates a trade distortion in favour of those uncapped countries that do not have an effective carbon price. Put bluntly, not pricing carbon properly on a comparative basis in China is equivalent to an export subsidy[8].

The problems for a top-down approach such as Kyoto are made worse by the fact that the impacts of climate change are not all bad, and vary considerably. Arctic countries, such as Russia and Canada, have much to gain from the resources that will become accessible once the ice has melted. For temperate zones, initial rises in temperature might mean that less energy will be needed for winter heating and, in some areas, could result in higher agricultural productivity.

Warming of the planet by more than 2 °C will change this cost–benefit equation, but if costs and benefits differ, getting a top-down agreement is made all the harder. For Russia and Canada, for example, the situation is very different from that in poorer tropical nations. No wonder both these Arctic countries were effectively in the 'reluctant camp' at the Durban COP, joining the United States on the outside.

Fortunately, there is a better way forward. Instead of taking a top-down approach that requires global agreement, climate-change policies can be constructed from the bottom up using three key building blocks: putting a tax on carbon consumption; switching from coal to gas as quickly as possible; and boosting spending on new energy technologies. The first building block is to recognize that carbon consumption is more important than carbon production, and so the carbon price should be based on consumption. This means pricing the consumed carbon irrespective of where it was produced.

TAXING CARBON TRADE

If the exporting and importing countries both have domestic carbon prices, then it does not matter what the base is. But if an exporter of carbon-intensive goods — such as China — does not price carbon at an appropriate level, there needs to be a border tax adjustment on imports. Although there are lots of practical issues, only a small number of carbon-intensive industries make up the bulk of carbon trade, so in practice, a targeted set of border taxes should do the job.

The neat consequence of this approach is that countries can take their own measures to address carbon without reducing their competitiveness, and it encourages the exporter to introduce its own carbon price to avoid the money going to the importer's government.

Proper carbon pricing also helps to encourage the second building block — the coal-to-gas switch. In the short term, this switch is perhaps the only way of slowing down the wall of coal that will come into world energy systems in the next decade.

Pricing carbon and getting out of coal provide the only serious prospects for having some effect on emissions in the short term. Further ahead, what matters is technology. None of the existing technologies is well placed to crack the climate-change problem — especially not the current generation of wind turbines and solar panels. Biomass and energy crops have similar constraints: there is not enough land, water or shallow sea to yield sufficient energy to meet the needs of a global population that is predicted to rise to nine billion by 2050 (ref. 9).

The good news is that there are many new technologies that might help to crack the problem — from next-generation solar to geothermal and even new nuclear technologies. This is the third and most important building block.

Although politicians have put all their efforts into Kyoto, and Europe has overwhelmingly invested in current renewables partly as a result, the prize has been ignored. It would be better if the focus shifted to what really matters — pricing carbon consumption, getting out of coal as fast as possible and investing heavily in future renewables and new energy technologies.

Little, if any, of this will be achieved through the Kyoto approach. At each COP, the now well-established process of green groups and environmental ministers gathering together to make declarations of intent, without much concrete action, is played out. Copenhagen in 2009 was supposed to demonstrate Europe's world leadership in crafting a new climate agreement. In fact, the United States and China agreed the weak Copenhagen Accord outside the conference framework, and without Europe. At Durban, even the prospect of serious action was put off until after 2020.

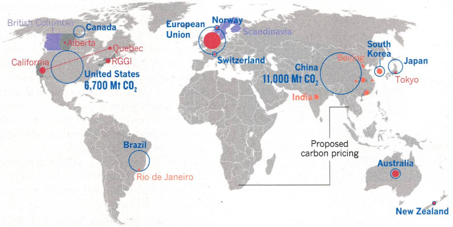
There is merit in world leaders meeting, not least because it puts climate change in the media spotlight. Yet the Kyoto conferences allow participants to be seen to be taking climate change seriously while actually doing very little. Climate-change policy cannot wait another decade. The net results of Copenhagen and Durban have been to make Kyoto largely irrelevant. It is probably worth going on with the talking, but the main chance lies in getting on with the bottom-up approach. ■

Dieter Helm *is professor of energy policy at the University of Oxford and fellow in economics at New College, University of Oxford, UK. His latest book is* The Carbon Crunch *(Yale Univ. Press, 2012).*
e-mail: dieter.helm@new.ox.ac.uk

1. Helm, D. in *The Carbon Crunch: How We're Getting Climate Change Wrong — and How to Fix It* 40–55 (Yale Univ. Press, 2012).
2. Helm, D., Smale, R. & Phillips, J. *Too Good to be True? The UK's Climate Change Record* (2007); available at http://go.nature.com/asmema.
3. European Commission. *20 20 by 2020: Europe's Climate Change Opportunity* (EC, 2008).
4. European Commission. *Off. J. Eur. Union* **L275,** 32–46 (2003).
5. BP. *BP Statistical Review of World Energy 2012* (BP, 2012).
6. International Energy Agency *Medium-Term Coal Market Report 2011* (IEA, 2011).
7. US Energy Information Administration. *Monthly Energy Review October 2012* (USEIA, 2012).
8. Helm, D., Hepburn, C. & Ruta, G. *Trade, Climate Change and the Political Game Theory of Border Carbon Adjustments* Grantham Research Institute on Climate Change and the Environment Working Paper No. 80. (2012); available at http://go.nature.com/4jop6c.
9. MacKay, D. J. C. *Sustainable Energy: Without the Hot Air* (UIT Cambridge, 2008).

CARBON TRADERS

Cities, regions and emerging economies are following Europe in setting up carbon-emissions trading schemes. The amount of carbon capped (filled circles) is shown to the same scale as national emissions volumes (blue circles).



○ Emissions (2010) ● Operating trading systems ● Legislated trading systems (by 2015) ● Carbon tax

RGGI, Regional Greenhouse-Gas Initiative. Mt CO_2, megatonnes of carbon dioxide equivalent (including deforestation).

SOURCE: A. PRAG, G. BRINER & C. HOOD MAKING MARKETS: UNPACKING DESIGN AND GOVERNANCE OF CARBON MARKET MECHANISMS (OECD/IEA, 2012).

Cap and trade finds new energy

An emerging coalition is implementing carbon trading despite political obstacles. It is rewriting the map of climate diplomacy, says **Michael Grubb**.

Fifteen years ago, the Kyoto Protocol presented a blueprint for a way to curb the increase of carbon dioxide in the atmosphere. One key measure, put forward by the administration of Bill Clinton, the US president at the time, was an enticing extension of free-market logic: establish emissions quotas to limit pollution, and trade them internationally. Let the market find the lowest-cost way to deliver the goal.

Cap and trade is one of two ways to put a price on pollution — taxation is the other. And emissions trading had been enacted successfully through the US Clean Air Act of 1990 to limit sulphur dioxide, a cause of acid rain.

The United States rammed a global cap-and-trade model into the Kyoto Protocol, allowing industrialized countries to meet their commitments with maximum flexibility, such as by buying emissions credits from projects in developing countries. The model was crucial in agreeing national emissions caps for Kyoto's first 'commitment period', from 2008–12.

That time has proved long enough for the grand plan of a globally negotiated, unified cap-and-trade system to unravel. The United States never ratified the protocol. Europe's flagship Emissions Trading Scheme (ETS) is in trouble. Yet, something unexpected is happening. Smaller carbon-pricing schemes are springing up across the world (see 'Carbon traders'). This renewed momentum could yet deliver an effective response.

ROCKY ROAD

Even though the idea of cap and trade originated in the United States, it fell to the European Union (EU) to implement it for CO_2 after US president George W. Bush repudiated the Kyoto Protocol in 2001. The EU's rapid establishment of a carbon cap and price covering CO_2 emissions from power generation and industry across 27 countries (plus some neighbours) was an astonishing achievement.

The EU ETS evolved in phases. Its first three years (2005–07) ended with the carbon price crashing to zero as industry — fearing a shortage of emissions allowances — instead achieved a surplus by over-complying with emissions cutbacks.

The second phase coincided with Kyoto's first commitment period. The European Commission won a historic legal and political battle to force EU member states to set emissions caps aligned with their national Kyoto commitments. To avoid another pricing roller coaster, the rules allowed for surpluses to be 'banked' for later use. In its first four years, the ETS is estimated to have cut CO_2 emissions by 40 million to 80 million tonnes a year on average, or 2–4% of the total capped[1].

Flush with apparent success, the EU advocated a global carbon market of similar schemes operating in the member countries of the Organisation for Economic Co-operation and Development (OECD) by 2015 and expanding to include all major emitters by 2020. With the US presidency of Barack Obama, who espoused cap and trade in his 2008 campaign, it seemed that such a vision was back on track. The US Congress began work on the Waxman–Markey energy bill, which had emissions trading as its centrepiece.

But events intervened and appetites for carbon trading waned. The credit crunch brought complex trading instruments into disrepute. The EU ETS saw scandals such as the theft of allowances from registries and fraud associated with complex tax treatments across borders.

In July 2010, the Waxman–Markey bill crashed in the Senate. The consensus became that the United States would not stomach carbon pricing — the public would never accept taxation, and the Senate had rejected the only alternative. Meanwhile, it became clear that the EU would over-achieve its second-phase Kyoto targets, owing to recession and progress on energy efficiency and renewable energy.

To give industry more time to plan, the third phase of the EU ETS was extended to 2020. This turned out to be a poisoned chalice. The surplus of emissions permits from phase two is now so big that it covers all of the cutbacks agreed up to 2020 — rendering the phase-three cap almost irrelevant.

The ETS carbon price has correspondingly slumped to around €7 (US$9) per tonne, less than one-third of that projected before Europe's recession.

This low price cannot incentivize investment in low-carbon technology, and it has devastated revenues that were expected to fund innovations such as carbon capture and storage. Instead, old plans for new and upgraded coal-fired power plants have been dusted off, predicated on the belief that the ETS surplus will sustain low carbon prices through 2020. Without changes, the future of the EU ETS — and its role in driving emissions reductions — looks bleak.

> **"Without changes, the future of the EU ETS in driving emissions reductions looks bleak."**

Many pundits rushed to declare emissions trading dead. Academics bashed targets and caps as a 'top-down' political process out of sync with the real world. The intelligentsia offered an alternative 'bottom-up' vision of localized mitigation efforts that would not rely on emissions caps, pricing or international negotiations[2].

These contributions underlined the importance of local politics and gave renewed attention to energy efficiency and innovation. But the alternative vision was also out of touch with reality, because it could not answer three fundamental questions.

First, major investments are made on the basis of economic returns. If there is no carbon price, why should investors decarbonize? The alternative is regulation directed against the grain of price signals — hardly appealing to conservative critics. Second, emissions targets set goals that can direct effort. Why would not specifying a target be more likely to spur action? Third, innovation requires years of nurturing, development, commercialization and growth, based on the prospect of profitable markets. Why should low-carbon technologies emerge faster without carbon goals and prices to reward and help to finance such innovation?

Some advocates of the new vision cite the decline in US emissions, despite the lack of a carbon price, as proof of concept[3]. But it does not follow that carbon trading and targets are irrelevant. High global oil and coal prices, economic slowdown, stronger energy-efficiency programmes and renewable energy policies have contributed to lower US emissions, as has an abundance of cheap shale gas. Both price and non-price factors are important. Cheap gas will reduce emissions if it displaces coal but not if it displaces renewable energy sources. The CO_2 savings achieved in recent years could be reversed if there is not a carbon price sufficient to ensure that gas remains cheaper than coal for generating power.

THE NEW RULES

In practice, events are challenging the 'top-down' and 'bottom-up' caricatures as well as the assumed leadership role of developed-country governments. With stalled national progress and a mix of concerns about climate change, energy security and stimulating investment and innovation, some states, cities and emerging economies are forging ahead with local carbon taxation and trading policies[4].

Asian economies have rushed to fill the void. In 2009, South Korea adopted the most environmentally oriented stimulus package of any country, and legislation for emissions trading passed its parliament in May 2012. Beginning in 2015, the scheme will set caps for facilities responsible for 60% of the country's emissions[5].

AFTER KYOTO
The legacy of a climate treaty
nature.com/kyoto

China's National Development and Reform Commission in July 2010 launched pilot 'low-carbon development zones' in five provinces and eight cities. Under the country's five-year plan for 2011–15, five cities (Beijing, Tianjin, Shanghai, Chongqing and Shenzhen) and two provinces (Guangdong and Hubei) will establish pilot emissions trading schemes. Adoption of a national cap-and-trade scheme by 2015 seems implausibly fast, but with the pace of Chinese developments, who knows.

India has developed the 'Perform Achieve Trade' system to implement energy-efficiency goals in three phases during 2012–20, for essentially the same sectors covered by the EU ETS.

> *"Smaller carbon-pricing schemes could yet deliver an effective response."*

In wealthy countries, several regional governments have pushed ahead. This month California held the first auctions for its CO_2 emissions trading system, with which Quebec is also affiliated. In 2008, British Columbia introduced the first major carbon tax since the Scandinavian countries did so 20 years earlier, with the price rising to Can$30 (US$30) per tonne in 2012. In 2011, Tokyo became the first city to adopt a municipal cap-and-trade system for carbon — an example emulated by Rio de Janeiro ahead of Rio+20, the 2012 United Nations Conference on Sustainable Development in Brazil[5].

Australia has clung to its domestic scheme, which came into force this summer as a three-year carbon tax that in 2015 will morph into a trading scheme linked with the EU ETS. The Australian government announced on 9 November that it intends to join the EU in signing on for a second period of the Kyoto Protocol.

By next year, about 10% of global emissions will be covered by a carbon price; by 2015, the Korean and Chinese pilot schemes will take this closer to 15% — and more plans are emerging. The effort will look nothing like the Kyoto vision or the EU's idea of an OECD-wide market. But it will be the first big step for an emergent 'coalition of the willing' that recognizes carbon pricing as integral to any credible strategy for sustainable and innovative low-carbon economic growth.

The dichotomy between 'bottom up' and 'top down' is false — it is like arguing over "whether a supertanker needs an engine or a captain", as Christiana Figueres, executive secretary of the United Nations Framework Convention on Climate Change, has put it (go.nature.com/mievx3). The 'engine' of emissions reductions is inevitably bottom up, and it must combine efforts on efficiency, pricing and innovation. But a top-down strategy, with national goals and international negotiations, is also essential to orient these efforts, address common problems and provide the international assistance required for global progress.

The Kyoto Protocol's grand plan is dead — so too is its antithesis. Carbon pricing is proliferating. Last year's United Nations conference in Durban, South Africa, launched negotiations for a global deal to be reached in 2015; this could reinforce, and help to link and orient, these emerging efforts.

NEXT STEPS

Three things need to happen. First, the United States must rejoin the global effort. Since Hurricane Sandy, New York City mayor Michael Bloomberg and New Jersey governor Chris Christie have opened the political window for renewed climate-change debate. The Obama administration needs to reinforce the message that the confluence of climate science and market economics demands a carbon price, whether by tax or trade.

Second, the EU needs a coherent suite of policies for efficiency, pricing and innovation. The ETS is a means to pricing and a route to financing efficiency and innovation, but it is not the whole game. The most urgent problem is to shift investment from new coal plants to low-carbon sources. There are grounds for removing the accumulated surplus of ETS emissions allowances; placing a minimum price on future auctions of allowances could deter coal plants and make the system more robust against future shocks[6].

Third, many developing countries in the United Nations take a negative negotiating position that is out of touch with their domestic progress and strategic interests. This is not a zero-sum game, and haggling over 'sharing the burden' is no solution. The emerging economies need to drop the blame game and use international negotiations to support their domestic efforts, so that clean, secure and sustainable energy overtakes carbon-intensive development worldwide as soon as possible. With these steps and initiatives elsewhere, a powerful coalition for effective emissions reductions could yet emerge and cement its progress in the global deal targeted for 2015. ■

Michael Grubb *is Chair of Energy and Climate Policy at the Cambridge Centre for Climate Change Mitigation Research, Cambridge CB3 9EP, UK.*
e-mail: mjg7@cam.ac.uk

1. Grubb, M. Laing, T., Sato, M. & Comberti, C. *Analyses of the Effectiveness of Trading in EU-ETS* (Climate Strategies, 2012).
2. Prins, G. & Rayner, S. *Nature* **449,** 973–975 (2007).
3. Lomborg, B. A fracking good story. *Project Syndicate* (13 September 2012).
4. Paterson, M. *Clim. Policy* **12,** 82–97 (2012).
5. Kosoy, A. & Guiogon, P. *State and Trends of the Carbon Market 2012* (World Bank, 2012).
6. Grubb, M. *Strengthening the EU ETS* (Climate Strategies, 2012).

ORJAN F. ELLINGVAG/CORBIS

Skyscrapers such as New York's Citigroup Center must contend with complex wind dynamics.

ENGINEERING

Turbulent genius

Allan McRobie enjoys a life of the audacious engineer who pioneered the windproofing of bridges and skyscrapers.

Siobhan Roberts' *Wind Wizard* is an unlikely gem, a biography of both a man and a field. It tells the story of Alan Davenport and the 50 years he spent creating the discipline of wind engineering to span the gaps between fluid dynamics, meteorology, structural engineering and architecture.

The book reveals the backstory to many of the world's more iconic structures. Here, for instance, are the World Trade Center; the Sears, CN and John Hancock Towers; the Citicorp Center; and a comparable compendium of epic bridges, all from the perspective of their ability to withstand windstorms. Early wind-tunnel tests of the World Trade Center revealed the need for more realistic wind modelling, spurring Davenport to establish the Boundary Layer Wind Tunnel at the University of Western Ontario in London, Canada. This facility was designed to replicate the turbulent conditions of the lower atmosphere.

WIND WIZARD
SIOBHAN ROBERTS

Wind Wizard: Alan G. Davenport and the Art of Wind Engineering
SIOBHAN ROBERTS
Princeton University Press: 2012. 288 pp. $29.95, £19.95

Such projects are massive investments, often of billions of dollars, with thousands of lives at risk when extreme winds hit. The physics is complex and uncertain, the mathematics intractable and the definitive experiment — building the full-scale structure and seeing what happens — cannot be done. It is a difficult problem, and the book describes how Davenport pieced together pragmatic theory and painstaking model testing to give rational, reliable predictions of performance.

Roberts charts how each challenge led to improvements in procedure and theory. For example, in her descriptions of the young Davenport's meetings with Leslie Robertson, the World Trade Center's structural engineer, as early as 1964, you detect both the creation of a prudent yet record-breaking design and the emergence of a field. Davenport replaced rudimentary rules of thumb for static pressures with a discipline. This tackled the complexities and uncertainties of the atmospheric boundary layer and the dynamic complications of wind flows such as galloping, vortex shedding, buffeting and wake buffeting. It was from that design process — which inevitably makes for poignant reading given the events of 11 September 2001 — that the Boundary Layer Wind Tunnel emerged. It went on to become central to all such studies.

Two of the projects studied in the tunnel show the potential for disaster posed by skyscrapers. In 1978, a phone call from an inquisitive student to the structural engineering firm behind the 59-storey Citicorp Center (now the Citigroup Center) — already built, and balancing on four huge columns high above mid-town Manhattan — prompted the shocking realization that the supporting calculations had omitted to take into account 'quartering' winds, which hit the building at 45 degrees.

This story, well-known in structural-engineering circles, represents one of the nightmare scenarios. Roberts captures the heart-thumping horror of the moment, and the parts played by Davenport and Robertson in the testing and emergency remedial action to strengthen the bracing that followed. Perhaps my

➲ NATURE.COM
For *Nature*'s city special, see:
nature.com/cities

only criticism of the book is that the student who telephoned is not named. I believe her to be Diane Hartley, who was then studying under David Billington at Princeton University — a surprising omission for that institution's own press.

The other difficult case is the John Hancock Tower in Boston. Its problems were more subtle, although equally alarming. Again, the issue involved wind forces from directions that had not been considered, and required the retrofitting of stiffening and dampers, at great expense, to make the structure safe. From now on, I shall refer students and professors alike to Roberts' clear account.

I did begin to wonder whether the ultimate outcome of Davenport's life-long effort was allowing financiers to inhabit lofty eyries without overly endangering the people below. But the last chapter focuses on his determined efforts at disaster mitigation for the vulnerable. For example, in the Caribbean, he has worked on hurricane-resistant houses and was involved in numerous international initiatives that worked on disaster mitigation at a human scale.

Roberts has written a largely equation-free book in which technical subtleties such as aeroelasticity and Davenport's statistical description of turbulent buffeting are set out clearly, engagingly and accurately. Her precise, vivid phrases, such as vortices "pushing and shoving the structure this way and that like a gang of bullies", will enliven my future lectures.

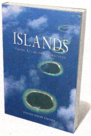
Two of the projects studied show the potential for disaster posed by skyscrapers.

Before opening the book, I had decided to look out for two potential pitfalls. First, would the book acknowledge the alternative to Davenport's statistical theory of buffeting — the rapid distortion theory developed by Julian Hunt? It does. Second, would the story of the famous 1940 Tacoma Narrows Bridge collapse in Washington state fall back on the lazy and inaccurate 'resonance' description that most physics textbooks adopt? It does not. Instead, Roberts gives faultless coverage of work by engineers Robert Scanlan and, more recently, Allan Larsen to explain the physics of what actually happened.

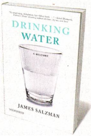
This is my field, but I learned much from Roberts' admirable book, and emerged with great respect for both Davenport and his chronicler. ■

Allan McRobie *is a Reader in Structural Engineering at Cambridge University, UK. e-mail: fam@eng.cam.ac.uk*

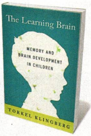
Books in brief



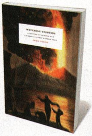
Islands: From Atlantis to Zanzibar
Steven Roger Fischer Reaktion Books *352 pp. £22 (2012)*
From Charles Darwin's moment in the Galapagos to the cultural efflorescence of Minoan Crete, islands are "crucibles and cradles" — laboratories, havens, touchstones. In this tour of their biology, geology and culture, linguist Steven Roger Fischer offers a taste of the million or so mini-biomes studding Earth's rivers, lakes and oceans. He is a brilliant guide, whether discussing the anti-cancer properties of the Madagascar periwinkle, Papua New Guinea's 500 languages, the imagined isle where Shakespeare's Prospero abjured his 'rough magic', or the very real threat climate change poses to many islands.



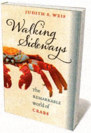
Drinking Water: A History
James Salzman Overlook *320 pp. $27.95 (2012)*
Potable water permeates humanity's past and is set to dominate its future. The United Nations estimates that by 2030, more than half of us will live in water-scarce areas. Environmental-policy specialist James Salzman goes with the flow in this absorbing chronicle of our complex relationship with H_2O. He negotiates multiple currents: the 'cures' ascribed to sacred waters; the ongoing struggle to eradicate microbes and dicey chemical compounds; urban waterworks and politicized availability; scarcity and bulk water transfers; and today's search for water — a quest awash with uncertainty.



The Learning Brain: Memory and Brain Development in Children
Torkel Klingberg Oxford Univ. Press *200 pp. $24.95 (2012)*
In this pragmatic treatise on how children learn, neuroscientist Torkel Klingberg homes in on working memory. Klingberg has marshalled swathes of research and pertinent case studies to show how gaps in this form of memory can lead to educational failure, and how training the young in tested techniques can help. Enriching his argument with findings — from the role of white matter to the corrosive effects of stress — he concludes that key pedagogic tools include 'memory training' to boost cognitive function, aerobic fitness, reduced anxiety and regular sleep.



Watching Vesuvius: A History of Science and Culture in Early Modern Italy
Sean Cocco Univ. Chicago Press *336 pp. $45 (2012)*
Historian Sean Cocco looks anew at Vesuvius to reveal how early responses to it shaped modern volcanology. Now monitored closely — as befits a looming risk to at least a million people — in Renaissance and Baroque Naples the volcano was just becoming a focal point for scientific appreciation. Cocco argues that a combination of the city's cultural traditions and the chain of eruptions that kicked off in 1631 helped to avert the early modern scientific eye from sky-gazing to the earthly wonders of geology.



Walking Sideways: The Remarkable World of Crabs
Judith S. Weis Comstock Publishing Associates *256 pp. $29.95 (2012)*
Stalked eyes, formidable claws, sidling gait: crabs are found around the globe and in environments ranging from deep-sea vents to bromeliad plants growing in trees. Biologist Judith Weis explores this crustacean cosmos with verve, touching on evolution, species, habitats, anatomy and functions, behaviour, ecology and fisheries. From the spotted orange Japanese spider crab (whose leg span can measure more than 3.5 metres) to the shell-swapping hermit crabs of Belize, this is a gripping overview of a remarkable family.

HISTORY

Dreaming of the bomb

Istvan Hargittai explores a life and work of Manhattan Project leader, physicist J. Robert Oppenheimer.

A towering yet enigmatic figure among theoretical physicists, J. Robert Oppenheimer directed the US laboratory in Los Alamos, New Mexico, that, between 1943 and 1945, built the first atomic bombs. He earned the label 'father of the atomic bomb' and worldwide fame, and features in numerous books. In the latest, *Inside the Centre*, Ray Monk — biographer of Bertrand Russell and Ludwig Wittgenstein — brings a philosopher's nuanced perception to Oppenheimer's life and work.

Oppenheimer grew up in a privileged upper-west-side Manhattan family, but felt burdened by being Jewish and "tried to pretend that he wasn't", in the words of his friend, the Nobel-prizewinning physicist Isidor Rabi. A lonely childhood was followed by a troubled youth; he even showed signs of destructive tendencies. Oppenheimer was trying, as he would all his life, to discover an identity and an avocation.

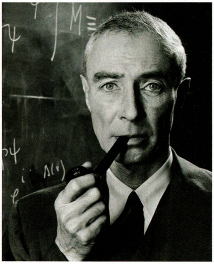
Oppenheimer followed the customary path of budding US scientists of the time, completing his education in Europe. In 1925, he joined Ernest Rutherford's Cavendish Laboratory in Cambridge, UK, where he was mentored by future Nobel prizewinner Patrick Blackett. Rumours persist of a bizarre incident in which Oppenheimer left an apple laced with a chemical — believed to be cyanide — on Blackett's desk. In any case, Oppenheimer was unhappy: he had little aptitude for experimental physics. Moving to Max Born's lab in Göttingen, Germany, a hotspot of theoretical physics, he became a top player.

In 1929, Oppenheimer returned to the United States for good. He worked at the California Institute of Technology in Pasadena and the University of California, Berkeley, building up an American school of theoretical physics. Soon, an influx of brilliant scientists fleeing the Nazi takeover in Europe arrived to bolster his efforts. Among the glowing successes were contributions to what later became known as the black-hole concept and astrophysics. By the time the field could contribute to the war effort, he and his colleagues were ready.

For a long time, the well-to-do Oppenheimer was oblivious to the economic difficulties around him and had little interest in world affairs. His political awakening in the mid-1930s occurred as a consequence of the hardship he observed during the Great Depression and the intensifying persecution of Jews in Germany. He was drawn to the Communist Party, although he always denied having been a card-carrying member.

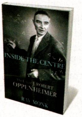
When nuclear fission was discovered in Germany in 1938, the Manhattan Project was initiated to develop an atomic weapon. Its final phase was bomb production — for which the Los Alamos Laboratory was created in 1943. This powerhouse drew in other Manhattan Project resources: brainpower from the Metallurgical Laboratory in Chicago; uranium-235 from Oak Ridge, Tennessee; and plutonium from Hanford, Washington. Oppenheimer, however, seemed an odd choice as leader, having never directed anything. What no one foresaw was his remarkable ability to inspire associates.

Oppenheimer never regretted his role in making the bombs. He saw their deployment against Japan as helping to end the Second World War quickly, saving millions of lives, despite having killed some 150,000 Japanese in Hiroshima and Nagasaki. In 1947, he declared that "physicists have known sin". Later, he clarified that he meant the sin of taking pride in their achievements rather than the sin of having caused destruction.


Robert Oppenheimer in 1958.

PHILIPPE HALSMAN/MAGNUM

Once involved with the Manhattan Project, Oppenheimer gradually dissociated himself from communism. However, even while directing Los Alamos, he was constantly being investigated by US security organs over his communist activities and connections. In his eagerness to demonstrate loyalty to his country, Monk reveals, Oppenheimer lied despicably about friends and former pupils. For example, he unjustly accused his gifted former student, Bernard Peters, who had participated in anti-Nazi street-fights in Germany, of being a dangerous Red.


Inside the Centre: The Life of J. Robert Oppenheimer
RAY MONK
Jonathan Cape: 2012. 832 pp. £30

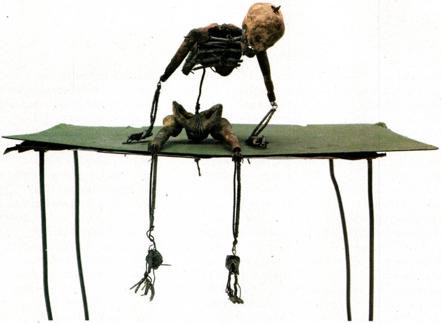
After the war, Oppenheimer was in great demand, and seen as a hero scientist. He chaired several committees, including the General Advisory Committee of the Atomic Energy Commission (AEC), which sometimes caused conflict of interest. For example, the Pentagon gave up the idea of the hydrogen bomb after Oppenheimer told them it was technically unfeasible. He then told the AEC that the Pentagon wasn't interested in developing the bomb. Spreading himself too thin also impaired his judgement: he humiliated others, made powerful enemies and hurt his chances of maintaining a leading role in government affairs, which he craved.

During the McCarthy era between 1950 and 1954, Oppenheimer's leftist past caught up with him. His concocted stories surfaced, and his only explanation was: "I was an idiot." Monk's presentation of the well-known story of the 'Oppenheimer hearing' before an AEC security panel is a highlight of the book.

Oppenheimer had the highest level of security clearance because of his sensitive position. By the time his clearance was about to expire, his loyalty and trustworthiness had been questioned by a number of people. The AEC set up a personal security board to decide on an extension and, in 1954, many scientists testified before it. The damaging testimony of nuclear physicist Edward Teller is often held responsible for Oppenheimer's downfall. The most relentless advocate for a US hydrogen bomb, Teller viewed Oppenheimer as an obstacle to his efforts. But the 'prosecution' had already destroyed Oppenheimer's veracity by the time Teller stepped into the witness stand. Teller's testimony ultimately harmed him more than it did Oppenheimer.

Oppenheimer was both a brilliant physicist and a poor politician; a sophisticated speaker and an inconsistent debater; an inspirational colleague and a disloyal friend. In this highly readable book, Monk makes great strides towards fully understanding the phenomenon that was J. Robert Oppenheimer. ■

Istvan Hargittai *is professor emeritus at Budapest University of Technology and Economics and author of* Judging Edward Teller.
e-mail: istvan.hargittai@gmail.com



JUNE LEAF *GENTLEMAN ON GREEN TABLE* (1999–2000)/EDWARD THORP GALLERY

THANATOLOGY

Beyond the grave

Death's multifarious faces in two London exhibitions exhilarate **Ewen Callaway**.

Some people hoard comic books, others sports cars, dolls or door knobs. Pretty much anything that can be culled, catalogued and curated is being amassed by someone, somewhere.

Richard Harris collects death. More than a decade ago, the former antique-print dealer and one-time anatomy student came across a series of *memento mori* prints, such as a late eighteenth-century engraving that depicts a half-man, half-skeleton digging his own grave. That print, along with several hundred other works that Harris has since acquired, is now on show at *Death*, a marvellous exhibition at London's Wellcome Collection.

"It's a universal subject," Harris said, as gallery workers put the finishing touches to the five-room, 300-piece exhibit. "We're all going to die."

The vast Wellcome Collection — comprising medical devices, texts and miscellanea — is an appropriate setting for Harris' collection, which shattered attendance records when it was shown at the Chicago Cultural Center in Illinois earlier this year. But it is safe to say that Henry Wellcome, the millionaire who amassed the bulk of the Collection, never owned anything quite like Jodie Carey's 4-metre-tall *In the Eyes of Others* (2009). One of the work's three chandeliers, which are made of hundreds of plaster-cast bones, illuminates the entrance.

➲ **NATURE.COM**
For more on art and death, see:
go.nature.com/9rcpjk

Bones abound at *Death*. Barthel Bruyn the Elder's sixteenth-century *Portrait of a Man/A Skull in a Niche* is a two-sided painting: nobleman on one side, a skull on the other. One wonders whether its original owner oriented it according to his health and mood. Contrast that work with the Argentine collective Mondongo's *The Skull Series* (2009) — a 2-by-2-metre skull made of miniature plastic books, buildings and a rubber duck, set against a backdrop that depicts *Pacman* video-game screens. 'Conversation piece' doesn't do it justice. Harris's own favourite is also the smallest: June Leaf's *Gentleman on Green Table* (1999–2000), a hunched skeletal form made of rusted tin, wire and screws. It is more evocative than any pile of calcium carbonate I've seen or held.

Yet there is a sense of osseous overload. By the time you get to Marcos Raya's imaginative series of Mexican portraits, inspired by Day of the Dead folk art, with skeletons superimposed on each family member, you have seen bones forged from brass, papier mâché and laser-cut paper.

The most powerful pieces tackle death directly. In a series of 51 etchings, the German expressionist Otto Dix depicts his time as a First World War artillery gunner. *Storm Trooper Advancing Under Gas* (1924) shows ghoulish gas-mask-clad soldiers emerging from a trench. A century earlier, Francisco Goya captured the horrors of the Peninsula War between France and Spain in a series of haunting etchings, *The Disasters of War* (1810–20). Corpses, a severed head and limbs dangle casually from a tree in *An heroic defeat! With dead men!*. These etchings, along with those of Jacques Callot from the end of the Thirty Years War, resemble war photojournalism in their matter-of-fact portrayal of brutality.

Death gets a more informative treatment across town, at the Museum of London's fascinating *Doctors, Dissection and Resurrection Men*. If real-estate tycoons are the primary beneficiaries of London's redevelopment boom, archaeologists come in a solid second; they often gain access to long-buried historical sites when new construction peels back a layer of the city's past. (*Nature* last year published the genome of the bacterium responsible for the Black Death, collected from bones excavated by Museum archaeologists in a fourteenth-century plague pit.)

In 2006, Museum of London archaeologists unveiled a nearly 200-year-old cemetery adjacent to the Royal London Hospital in Whitechapel. Their excavation revealed graves containing jumbles of bones from many people, as well as the odd turtle and cow, showing signs of amputation and dissection.

With the surgical profession on the rise in the early nineteenth century, medical students at private anatomy colleges needed cadavers for study. Legitimate sources — executed prisoners — were scarce, so anatomy schools sought the services of the grave robbers known as 'resurrection men'. The text displays at the exhibition are enriched with a range of journal clippings, letters and other primary sources.

Resurrection men could earn a handsome salary for digging up one body, and some even turned to murder. The slaying of an Italian boy and the trial of his murderers is told in detail, through video and documents. The killers were convicted and executed, and their bodies given over for dissection. The episode sparked widespread public revulsion and led to the passage of the Anatomy Act of 1832. This allowed the "unclaimed and friendless" bodies of indigent hospital patients to be used for dissection.

Among a sometimes intrusive welter of videos and interactive displays is the thoughtful *The Body Beyond Death* (2012), in which Londoners give their views on mortality. One, a young woman in a headscarf, explains why she would not want to donate her organs — poignantly showing how, for some, the dead are much more than a pile of bones. ■

Ewen Callaway *is a reporter for* Nature.

Correspondence

Lichens under threat from ash dieback

The fungal pathogen *Chalara fraxinea* is killing ash trees (*Fraxinus excelsior*) throughout Europe. Also potentially under threat is the large diversity of lichens that these ash trees support.

Using data from the UK National Biodiversity Network (www.nbn.org.uk), we found 536 lichen species (corresponding to some 30% of UK lichens) that occur on ash. Of these, 84 are categorized as under threat in Britain using International Union for Conservation of Nature standards.

For at least six of these threatened species, more than half of the records in the database are for specimens found on ash trees. This includes *Fuscopannaria ignobilis*, a lichen that receives the highest UK legislative protection status under Schedule 8 of the 1981 Wildlife and Countryside Act, and *Wadeana dendrographa*, for which the United Kingdom has international conservation responsibility.

Ash, along with non-native tree species such as sycamore (*Acer pseudoplatanus*), provided an alternative host for lichens affected by the catastrophic decline of elm trees during the 1970s. If the UK ash population succumbs to dieback, the rescue effect for lichens is a consideration that should influence landscape management of non-native trees.

Christopher J. Ellis, Brian J. Coppins, Peter M. Hollingsworth *Royal Botanic Garden Edinburgh, UK.*
c.ellis@rbge.org.uk

NIH chimps: don't sell sanctuary short

You quote Kathy Hudson, deputy director for science, outreach and policy at the National Institutes of Health (NIH), as saying: "In a perfect world, we would absolutely like to move all of the [retiring laboratory] chimps directly to Chimp Haven [sanctuary]" (*Nature* **491**, 18; 2012). This fails to acknowledge the NIH's legal mandate to do just that.

Since the passage of the Chimpanzee Health Improvement Maintenance and Protection (CHIMP) Act 12 years ago, the NIH has done less than right by the law and by its research chimps in allowing labs to make discretionary decisions about the animals' retirement. It claims not to have sufficient funding to provide housing for retired chimps at the federal sanctuary, while financing the expansion of labs to accommodate chimps.

The CHIMP Act obliges the NIH to provide lifetime care for retired chimps. The agency's fiscal-reserve cap does not restrict it from finding more funds to fulfil its mandate, or limit its responsibility to do so. Meanwhile, the federal sanctuary should not be going short.

Theodora Capaldo *New England Anti-Vivisection Society, Boston, Massachusetts, USA.*
theodoracapaldo@neavs.org

NIH chimps: use existing facilities

Louisiana's Chimp Haven sanctuary for retired laboratory chimpanzees is requesting a further US$2.55 million from the National Institutes of Health (NIH) to construct housing for 110 chimps about to retire from the New Iberia Research Center, part of the University of Louisiana at Lafayette (*Nature* **491**, 18; 2012). However, the sanctuary has almost hit the NIH funding cap of $30 million, and other facilities are already available at the Texas Biomedical Research Institute's Southwest National Primate Research Center (SNPRC), of which I am director.

The housing at the SNPRC is cost-effective, high quality and similar to some housing at Chimp Haven. The SNPRC also provides extensive enrichment programmes for the animals.

Besides housing, the NIH needs to consider the pressing health-care needs of ageing chimps and the capacity of institutions to meet them.

The SNPRC's medical capabilities are state-of-the-art. For example, four veterinarians are employed for 141 chimps, compared with one for 130 chimps at Chimp Haven. Veterinarians at the SNPRC have 90 years' combined experience working with chimps. And the SNPRC has an on-site pathology lab equipped for testing within minutes of a medical emergency, a facility not available at Chimp Haven.

Given that the vacant facilities at the SNPRC were partly funded by a $1.5-million grant from the NIH, the proposed allocation of scarce federal research dollars to duplicate them at Chimp Haven seems wasteful. The Congressional Budget Office has estimated that the cost to the taxpayer of transferring the New Iberia animals and 330 other NIH-owned chimps to Chimp Haven would be $56 million over the next 5 years alone.

John L. VandeBerg *Texas Biomedical Research Institute and Southwest National Primate Research Center, San Antonio, Texas, USA.*
jlv@txbiomedgenetics.org

NIH chimps: Texas lab is not a sanctuary

As a physician who provided testimony for the US Institute of Medicine report that found chimpanzee experiments to be scientifically unnecessary, I am thrilled that the National Institutes of Health has declared some 110 of its chimps ineligible for research (*Nature* **491**, 18; 2012).

But sending any of these chimps to the Texas Biomedical Research Institute in San Antonio, even for the short term, is no retirement. Laboratories are designed to facilitate research, not to provide high-quality, long-term care for chimps, which are cognitively, socially and emotionally complex animals.

Furthermore, the Texas Biomedical Research Institute was fined more than US$25,000 by the US Department of Agriculture in December 2011 for violations of the Animal Welfare Act after three animals escaped from their cages (see go.nature.com/dnzsqa).

John Pippin *Physicians Committee for Responsible Medicine, Washington DC, USA.*
jpippin@pcrm.org

Cuts endanger young scientists in Europe

The Young Academy of Sweden has joined forces with the Young Academies of Germany, the Netherlands and Denmark to urge the leaders of the European Union to invest more, not less, in science in their upcoming budget (see go.nature.com/ymjole). Short-term savings would have long-term costs and weaken Europe's future scientific standing.

The academies' members are especially concerned that cuts could target the European Research Council (ERC), which has emerged as a funding model for junior researchers in high-profile international science.

The ERC provides these young scientists with the funding to develop their own lines of research, rather than relying on the patronage of a senior colleague. This strategy encourages promising scientists to stay in Europe and others to return from abroad. It also attracts research talent from around the world.

Christian Broberger, Anna Sjöström Douagi *The Young Academy of Sweden, Stockholm, Sweden.*
asd@sua.kva.se

RESEARCH

NEW ONLINE

Papers published this week at Nature.com/nature

PERSPECTIVE Just what is going on in the middle and upper stratosphere? **p.692**

ARTICLE Bread wheat supplies one-fifth of our calories, and here's its genome sequence **p.705**

LETTER An over-massive black hole provides 59% of the mass of a compact galaxy **p.729**

PUBLISHED 25 NOVEMBER 2012

CELL BIOLOGY

Cohesin ring break part of DNA repair

During DNA repair, the ends of a double-stranded break are resected by a nuclease to leave single-stranded tails. The tails pair with homologous sequences, usually within the sister chromatid. This preference is enforced by cohesin, a ring-like complex that holds the sister chromatids in close proximity. Alexandra McAleenan *et al.* now show that cohesin recruited to the site of DNA damage is cleaved by separase, and that cohesin dissociation is necessary to allow resection to occur. They speculate that if cohesin dissociation is conserved in mammals, therapies that rely on the generation of DNA lesions could be improved by simultaneous inhibition of the cohesin cycle.

Post-replicative repair involves separase-dependent removal of the kleisin subunit of cohesin
A McAleenan *et al.*
doi:10.1038/nature11630

NEUROSCIENCE

Ventral tegmental area GABA projections pause accumbal cholinergic interneurons to enhance associative learning
M T C Brown *et al.*
doi:10.1038/nature11657

MOLECULAR BIOLOGY

Coordinated control of replication and transcription by a SAPK protects genomic integrity
A Duch *et al.*
doi:10.1038/nature11675

Recombination-restarted replication makes inverted chromosome fusions at inverted repeats
K'I Mizuno *et al.*
doi:10.1038/nature11676

CANCER

Restriction of intestinal stem cell expansion and the regenerative response by YAP
E R Barry *et al.*
doi:10.1038/nature11693

PUBLISHED 28 NOVEMBER 2012

GENETICS

Complex algal genomes revealed

John Archibald and colleagues present the sequences of the nuclear genomes of two algae of remarkable complexity, *Guillardia theta* and *Bigelowiella natans*. These are transitional forms in the endosymbiotic acquisition of photosynthesis by engulfment of eukaryotic algae and they have four genomes: mitochondrial and plastid (chloroplast); nuclear, of host origin; and miniaturized 'nucleomorph', of endosymbiotic origin. Analyses reveal extensive genetic and biochemical mosaicism and point to mechanisms that may explain nucleomorph persistence.

Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs
B A Curtis *et al.*
doi:10.1038/nature11681

Algae's complex origins
S B Gould
doi:10.1038/nature11759

MATERIALS SCIENCE

Tunable nanowires synthesized

Nanowires hold promise for a variety of applications in electronics, energy and biomedicine. However a major hurdle is the large-scale production of high-quality nanowires. Lars Samuelson and colleagues have developed a low-cost aerosol-based way to synthesize gallium arsenide nanowires with throughput that is substantially better than that achieved by conventional methods. The method produces high-quality nanowires with tunable dimensions, good optical properties and spectral uniformity.

Continuous gas-phase synthesis of nanowires with tunable properties
M Heurlin *et al.*
doi:10.1038/nature11652

GENETICS

Genetic variation quantified

The National Institutes of Health Heart, Lung and Blood Institute's Exome Sequencing Project has sequenced the exomes of more than 6,500 individuals of European and African American ancestry. Using these data, Wenqing Fu *et al.* estimate that about 73% of all protein-coding single nucleotide variants arose in the past 5,000–10,000 years. Around 86% of the variants predicted to be deleterious arose within the same time frame. The data suggest that the increased mutational capacity of recent human populations has influenced the burden of Mendelian disorders, but it is also likely to promote beneficial genetic changes that will be selected for in generations to come. More practically, the results will be useful in prioritizing potential disease-causing variants in gene-mapping studies.

Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants
W Fu *et al.*
doi:10.1038/nature11690

CELL BIOLOGY

The NAD-dependent deacetylase SIRT2 is required for programmed necrosis
N Narayan *et al.*
doi:10.1038/nature11700

Death by deacetylation
W Zhou & J Yuan
doi:10.1038/nature11761

BIOCHEMISTRY

Bypass of a protein barrier by a replicative DNA helicase
H Yardimci *et al.*
doi:10.1038/nature11730

Molecular hurdles cleared with ease
M A Trakselis & B W Graham
doi:10.1038/nature11758

ORGANIC CHEMISTRY

Practical and innate carbon–hydrogen functionalization of heterocycles
Y Fujiwara *et al.*
doi:10.1038/nature11680

Toolkit of reagents to aid drug discovery
W J Pitts
doi:10.1038/nature11760

CANCER

COUP-TFII inhibits TGF-β-induced growth barrier to promote prostate tumorigenesis
J Qin *et al.*
doi:10.1038/nature11674

BRIEF COMMUNICATIONS ARISING

Arising from 'A signature of cosmic-ray increase in AD *774–775 from tree rings in Japan' by F Miyake* et al. Nature ***486,*** *240–242 (2012)*

Causes of an AD 774–775 ^{14}C increase
A L Melott & B C Thomas
doi:10.1038/nature11695

COMPUTATIONAL MATERIALS SCIENCE

Substitution with vision

A method has been developed for predicting the stability and elasticity of certain alloys for millions of atomic configurations of the materials. This approach should help to identify materials with optimized properties. SEE LETTER P.740

GUS L. W. HART

Early civilizations began alloying copper with arsenic, tin or zinc nearly 6,000 years ago[1], ushering in the Bronze Age. Even when more-abundant iron became a mainstay during the Iron Age, substitutional alloys of copper — in which some of the copper atoms were replaced with atoms of a different metal — were superior materials. So Roman foot soldiers were equipped with wrought-iron weapons, whereas Roman officers had swords made of bronze. Similarly, at the turn of the twentieth century, French scientists developed a steel that substituted some iron with vanadium. This alloy, which had three times the tensile strength of competing steels, became an essential ingredient of the venerable Ford Model T (and of early French luxury cars).

Materials substitution — the replacement of some of the atoms of a material with those of another (Fig. 1) — continues to be a key strategy for developing the materials of tomorrow, but predicting the properties of new alloys is remarkably difficult. On page 740 of this issue, Maisel *et al.*[2] report a method for calculating from first principles both the elasticity and the thermodynamic stability of alloys*.

The difficulty of developing improved materials is a bottleneck — perhaps the main bottleneck — to advances in new technologies. In 2011, to increase the pace of materials development, and to leverage impressive advances in computational materials science, US President Barack Obama announced the Materials Genome Initiative, a project that aims to create an infrastructure of informatics and experimental tools for materials development in the United States[3]. Materials substitution is central to this initiative and has also been specifically referred to in recent calls for research proposals from US federal funding agencies.

Many computational approaches have been developed to take advantage of the materials-substitution strategy, but Maisel and colleagues' report brings something new to such efforts. Their work is noteworthy because it demonstrates a clear correlation between the thermodynamic and elastic properties of alloys known as face-centred-cubic intermetallics, and because their approach can easily be applied to other alloy types. It therefore not only provides a fundamental understanding of the physical properties of materials, but also opens up opportunities for materials engineering. Using the authors' method, it may be possible to tune the elastic stiffness of alloys using materials substitution. For example, alloys could be softened to make compounds for orthopaedic implants that integrate well with bone, in order to avoid the difficulties that arise when bone and implant materials have disparate elastic properties.

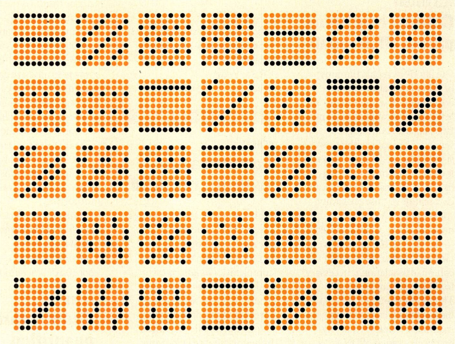


Figure 1 | Atomic configurations. Substitutional alloys are formed when a fraction of the atoms of a metal are replaced with different atoms. Several atomic configurations of possible alloys are depicted here for a hypothetical case in which the atoms of the main metal (orange) form a square lattice. Substituted atoms are shown in black. Maisel *et al.*[2] report a computational technique that allows rapid prediction of both the elastic stiffness and the thermodynamic stability of different atomic configurations of certain alloys.

The authors' technique extends a computational methodology known as cluster expansion that is often used to calculate the properties of substitutional alloys. Cluster expansion involves two basic steps: first, calculate the target property for a number of different atomic arrangements using quantum mechanics; and second, map this information onto a simple model that accounts for the effects of atomic substitutions. In this way, one essentially 'trains' a computational model, which is then used to calculate the target property for any atomic configuration — instantly and with quantum-mechanical accuracy. Because computation of the target quantity is so efficient, cluster expansion can be used to simulate thermodynamic and kinetic properties of atomic ensembles from first principles, and to screen hundreds of millions of atomic configurations for a specific property.

Maisel and colleagues report two major advances in cluster expansion. First, they have expanded its use to calculate multiple properties in a single model; and second, they have used it to identify a specific relationship between those properties. Specifically, they combined two cluster expansions to predict both the thermodynamic stability and the mechanical stiffness of any atomic

---

*This article and the paper under discussion[2] were published online on 21 November 2012.

configuration in their target alloys. This revealed that the more thermodynamically stable the configuration, the stiffer the resulting material.

Many researchers use computation to identify atomic structures of a given material that have desirable properties, often with little regard for whether those structures are feasible to make (thermodynamically stable). Maisel and co-workers' approach shows that, at least in the case of elastic stiffness, hunting for metastable structures that have better properties than stable structures — whether known or predicted — is essentially futile, and that researchers should focus on other materials instead. That said, being able to predict both the stability and another target property of a material will allow scientists to efficiently scan through sets of hypothetical materials and 'see' promising candidates, lending vision to an established computational approach. It is in this discovery mode that Maisel and colleagues' work could contribute greatly to efforts such as the Materials Genome Initiative.

It remains to be seen how many other materials' properties will be studied using the new approach. In principle, any property that directly depends on atomic configuration is within reach, but many properties of relevance to engineering are still difficult to compute in practice. Furthermore, some of the most important properties of materials depend not only on the atomic configuration of a fixed lattice, but also on microstructural elements — such as boundaries between microscopic crystals (grains), grain sizes and extended crystal defects. These remain beyond the reach of quantum-mechanical calculations.

Still, high-throughput approaches[4,5] for sifting through thousands, or tens of thousands, of candidate materials are poised to make a substantial contribution to society's needs by generating large databases of information that will be of use to researchers[6,7]. These databases will be more effective if the information they contain about physical properties is used to build computational models that, in turn, could search for thermodynamically stable materials to meet a particular need[8]. Maisel and colleagues' work, coupled with automated model-building methods[9], might help us to achieve that goal. ■

Gus L. W. Hart *is in the Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA.*
e-mail: gus.hart@byu.edu

1. Sass, S. L. *The Substance of Civilization* (Arcade, 1998).
2. Maisel, S. B., Höfler, M. & Müller, S. *Nature* **491,** 740–743 (2012).
3. www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf
4. Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. *Phys. Rev. Lett.* **91,** 135503 (2003).
5. Curtarolo, S. *et al. Comput. Mater. Sci.* **58,** 218–226 (2012).
6. Curtarolo, S. *et al. Comput. Mater. Sci.* **58,** 227–235 (2012).
7. Jain, A. *et al. Comput. Mater. Sci.* **50,** 2295–2310 (2011).
8. Yang, K., Setyawan, W., Wang, S., Buongiorno Nardelli, M. & Curtarolo, S. *Nature Mater.* **11,** 614–619 (2012).
9. Nelson, L. J., Zhou, F., Hart, G. L. W. & Ozoliņš, V. preprint at http://arxiv.org/abs/1208.0030 (2012).

PLANT ECOLOGY

Forests on the brink

An analysis of the physiological vulnerability of different trees to drought shows that forests around the globe are at equally high risk of succumbing to increases in drought conditions. SEE LETTER P.752

BETTINA M. J. ENGELBRECHT

Water is the most limiting factor for ecosystem diversity and productivity worldwide. But the global climate is changing, and both warming and shifts in rainfall patterns are projected, which will leave large areas of the planet with less rain and a higher likelihood of extreme drought events[1,2]. These changes will almost certainly affect forests, which cover more than 30% of the world's land surface. Understanding these effects is imperative: forests play an integral part in carbon and water cycles, they provide timber and other products, and they are home to a vast diversity of plants, animals and microorganisms. But forests occur in a wide range of climatic conditions, so it is a challenge to predict how the vulnerability of trees to changes in water availability compares between different biomes. In this issue, Choat *et al.*[3] (page 752) use a combination of physiological measurements of the vulnerability of trees to drought and of the drought stress they actually experience in their natural habitats to show that forests worldwide are at high risk*.

We might expect that trees in forests currently exposed to seasonal or multi-annual droughts, such as in 'Mediterranean-type' systems, are already well adapted and will therefore suffer less from an increase in drought conditions than trees in wet forests. Conversely, but equally reasonably, we could predict that trees in dry areas are already at their physiological limits and would therefore be more vulnerable to increased drought than trees in wet forests. To investigate these questions, Choat and colleagues compared the vulnerability of the tree water-transport system to drought in different species worldwide.

In plants, water is transported through a tubing system, a tissue called xylem that is made up of a multitude of conduits. Loss of water vapour (transpiration) through stomata (pores) in the plants' leaves generates suction that pulls water in the xylem from the soil through the roots and stem to the leaves — much like sucking water through a straw. On its way, the water provides crucial services to the plant: it is the medium for metabolic reactions, it transports nutrients and other substances, and it provides stability. However, the powerful suction that pulls water through the xylem brings with it the risk of pulling air through small holes, called pit pores, in the sides of the conduits. These air bubbles can block the xylem and impair water transport, just like sucking air into a broken straw. This process is called xylem embolism, and the higher the suction in the conduit, the more embolism occurs.

The link between this physiology and drought conditions comes from the fact that suction increases with increasing transpiration and/or decreasing water availability in the soil. Plants can regulate their stomata to delay the increase in suction, but if water is not replenished, more and more conduits will become clogged, leading to hydraulic failure and the eventual death of the plant. However, different plant species have different xylem structures, so the vulnerability of a plant's xylem conduits to embolism, and therefore its ability to tolerate drought, are variable.

The authors compiled data on the xylem vulnerability of 480 tree species from 183 sites worldwide, covering the broad range of climatic conditions in which forests occur. They included both angiosperms (flowering trees, such as oak and maple) and gymnosperms (such as pine and cedar), which vary substantially in their xylem structure. Wherever the data were available, they also included the maximum suction occurring in the trees in their natural habitats. Combining these data enabled Choat *et al.* to explore how the suction that induces hydraulic failure in a given species compares with the suction that it actually experiences. If these values are close together, this represents a small 'safety margin' with respect to hydraulic failure and indicates that the species is at risk; if they are far apart, the species is likely to be able to withstand more intense drought conditions.

The data show that, as expected, trees growing in more arid conditions around the globe

---

*This article and the paper under discussion[3] were published online on 21 November 2012.

50 Years Ago

It seems to be generally agreed that the standard of self-expression in spoken and written English among sixth-form and undergraduate scientists and technologists is low. Various causes have been blamed ... but in all the diagnoses and cures I have seen so far, all the emphasis has been on past failures by English experts and future remedies to be administered by other English experts. It is not my intention to dissociate English teachers from the problem altogether ... but I want to suggest that scientists and technologists themselves must take most of the responsibility for the low standards of self-expression in their professions, and that a major change of outlook on their part is the only thing that can bring a substantial improvement in the situation.
From *Nature* 1 December 1962

100 Years Ago

Biologische und morphologische Untersuchungen über Wasser- und Sumpfgewächse. By Prof. H. Glück — Prof. Glück has produced a portentous volume on the riparian flora, forming the third instalment of his work on water and swamp plants. Frankly, we do not find justification for the 600 or more pages of his book, and we fancy most readers who have been in the habit of using their eyes when observing or collecting plants will find but little to reward them for the trouble of its perusal ... No doubt a work of this kind possesses some value, but, as it appears to us, it excellently illustrates the truth of the saying that the secret of dullness lies in the attempt to write all one knows. Prof. Glück gives the impression (perhaps unjustly) that he has written all he knows about his subject, and certainly he has jotted down a good deal that is already very familiar to others.
From *Nature* 28 November 1912

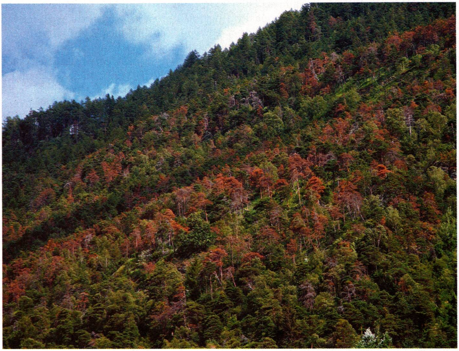

B. WERMELINGER, WSL

Figure 1 | Thirsty trees. Reports of drought-induced forest die-off[4], such as that in Switzerland in 1999 shown here, have increased in recent decades, suggesting that climate change is already having an impact on tree health in many locations. Choat and colleagues' study[3] of trees across the globe suggests that they are at high risk from even small increases in drought intensity.

are better at withstanding xylem embolism. The exciting finding, however, is that angiosperm trees in all forest biomes have converged on a risky strategy, operating at extremely narrow safety margins. This implies that these trees are already, under current conditions, on the verge of injurious levels of water availability, and that even a minor increase in drought intensity will induce levels of xylem embolism that will impair growth and lead to tree death.

The suggestion that all forests are on the brink of succumbing to drought, and may already be responding to climate change, is supported by observations of increased drought-induced forest die-offs and tree mortality in many ecosystems[4] (Fig. 1). For gymnosperms, Choat *et al.* found wider safety margins, suggesting that these trees may have a higher tolerance to increased drought. However, even these trees are threatened by hydraulic failure, as recent regional die-offs of pines show[4]. Taken together, these studies sound a warning bell that we can expect to see forest diebacks become more widespread, more frequent and more severe — and that no forests are immune. The ramifications of this scenario are diverse and, in many respects, dire: forest mortality will be accompanied by changes in species composition, changes in ecosystem function and losses of services and biodiversity[4].

Advancing our knowledge of organismal responses to factors such as drought and temperature is essential to improving predictions of the consequences of climate change[5,6]. Through their meta-analysis of the global distribution of xylem vulnerability, Choat *et al.* have dramatically increased our understanding of the comparative vulnerability of forests. Nevertheless, the mechanisms that actually lead to drought-induced tree mortality still remain elusive; in fact, it is known that some species can survive complete hydraulic failure for extended periods of time[7]. Although many studies have assessed the response of plants to experimentally manipulated precipitation and/or temperature[8], the results of these studies do not lend themselves to comparisons of drought responses across biomes, because of differences in treatments and in the resulting drought intensities. A coordinated network of standardized experiments is needed to further advance understanding of climate-change responses in ecosystems worldwide.

Our ability to forecast the consequences of drought for forests is also limited by the high regional uncertainty in current models for rainfall and drought prediction, for both long-term trends and extreme events[1,2]. A fundamental lesson from Choat and colleagues' study is that even small changes in drought intensity can be expected to lead to mortality in forests all over the world. This only highlights the urgent need for climate models that return more-confident predictions. ■

Bettina M. J. Engelbrecht *is at the Bayreuth Center of Ecology and Environmental Science, Department of Plant Ecology, University of Bayreuth, 95440 Bayreuth, Germany, and at*

the Smithsonian Tropical Research Institute, Panama.
e-mail: bettina.engelbrecht@uni-bayreuth.de

1. Parry, M. L. *et al.* (eds) *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007).
2. Field, C. B. *et al.* (eds) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2012).
3. Choat, B. *et al. Nature* **491,** 752–755 (2012).
4. Allen, C. D. *et al. Forest Ecol. Mgmt* **259,** 660–684 (2010).
5. Svenning, J.-C. & Condit, R. *Science* **322,** 206–207 (2008).
6. Craine, J. M. *et al. Nature Clim. Change* http://dx.doi.org/10.1038/nclimate1634 (2012).
7. McDowell, N. G. *Plant Physiol.* **155,** 1051–1059 (2011).
8. Wu, Z., Dijkstra, P., Koch, G. W., Penuelas, J. & Hungate, B. A. *Glob. Change Biol.* **17,** 927–942 (2011).

EARTH SCIENCE

Magma chambers on a slow burner

An assessment of crystallization processes occurring in magma chambers in the ocean floor finds an unexpected enrichment in trace elements, reviving an old theory of the cycling of magma in these chambers. **See Article p.698**

ALBRECHT W. HOFMANN

The world's ocean basins are constantly being regenerated by an 80,000-kilometre-long volcanic system of mid-ocean ridges, where Earth's mantle melts to form magma that eventually produces the basaltic floor of the oceans. The composition of ocean-floor basalts is one of the main sources of information about Earth's deeper interior. On page 698 of this issue, O'Neill and Jenner[1] re-examine the chemical compositions of basaltic lavas from this global magmatic system. They find new, and remarkably systematic, chemical relationships between the concentrations of 'incompatible' trace elements (so named because they are largely excluded from magmatic crystals) and that of magnesium oxide (MgO).

As expected, the content of incompatible elements increases in the basaltic-liquid component (the melt) of magmas, because MgO-bearing crystals precipitate in sub-oceanic magma chambers (reservoirs), causing the MgO content of the liquid to decrease. But O'Neill and Jenner show that the observed incompatible-element increase is much greater than conventional crystallization processes can explain. Their proposed solution to this dilemma would require a revision in the way geochemists calculate the composition of parental magmas entering magma chambers, and therefore also the composition of the mantle rocks from which these magmas are derived.

When basaltic lava comes into contact with cold sea water, it is chilled into glass. Geochemists like to analyse such glasses because they preserve the chemical composition of the lava particularly well, and they can thus tell the researchers much about the composition of the underlying mantle in which the melt forms. However, this view of the mantle is blurred because there are several intervening stages between melt formation and the eruption of lava. These are: partial melting of the mantle at depth (greater than about 30 km); extraction of the melt from the partially molten mush; its emplacement in shallow magma chambers; the formation and settling out of magmatic crystals in these chambers; and, finally, eruption of the remaining liquid on the ocean floor.

Two fundamentally opposing views of the mantle composition inferred from these glasses have prevailed over the past 40 years. One holds that the mantle has an essentially uniform composition, and that the compositional variability of the erupted basaltic lavas is produced primarily by processes occurring in the shallow magma chambers. The other view holds that magma-chamber processes have only minor effects on the erupted lavas that can be easily corrected for, and that the variations in lava composition mainly reflect differences in the composition of the mantle source and in the specifics of the melting process.

This latter view has gradually gained the upper hand, because much of the observed chemical variability of the lavas correlates with variations in the isotopic composition of the elements strontium, neodymium, hafnium and lead. These elements are the products of very slow radioactive decay, and therefore accumulate only during long residence times in the mantle. The observed differences in isotopic composition can therefore not be produced in short-lived magma chambers, but require long-term differences in parent–daughter ratios in the (mantle) source of the melts.

A crucial requirement when going backward from observed compositions of erupted basalts to their mantle sources is to evaluate the effects of partial crystallization and loss of the crystals in magma chambers. This is widely assumed to involve 'fractional crystallization', whereby newly formed crystals are immediately removed from chemical interaction with the liquid. Laboratory experiments have shown that the crystallization process in ocean-ridge magma chambers invariably involves the magnesium-bearing mineral olivine. The net effect of this is that the MgO content of the liquid progressively decreases as freshly crystallized olivine is removed from the liquid, whereas there is an increase in the contents of incompatible trace elements (such as barium, thorium and neodymium) because they are excluded from the crystals.

This was thought to be well understood — until O'Neill and Jenner plotted the incompatible-element contents against MgO for two recently assembled global data sets[2,3]. They found excellent linear correlations (with the expected negative slopes) between incompatible-element and MgO contents. However, they were startled to find that these slopes are consistently greater than the maximum allowed from fractional-crystallization theory.

If fractional crystallization does not explain this effect, what process does? One possibility is that lavas that have higher incompatible-element contents start out with systematically lower parental MgO contents. But that would mean that the sources of these magmas could not contain olivine, even though this is the most common of all upper-mantle minerals. Nevertheless, it has been proposed[4] that some mid-ocean-ridge basalts are mixtures of liquids formed from peridotite, the 'standard' olivine-bearing mantle rock, and other liquids formed from eclogite or pyroxenite, which are olivine-free rocks that form from subducted, recycled oceanic basalts. Melts from such recycled basalts should also have a higher-than-normal content of incompatible elements and a lower-than-normal MgO content. Such recycled basalts should also have different isotopic compositions of neodymium, for example. However, the expected correlations between neodymium isotopes and MgO have not been documented for any global set of ocean-ridge basalts.

As a way out of the dilemma, O'Neill and Jenner revive and generalize a model that was originally proposed by O'Hara[5] and later modified by Albarède[6], but which has been mostly forgotten. This model envisages a magma chamber that is periodically refilled with fresh parental liquid from below. The fresh liquid mixes with the pre-existing liquid, and the mixture is tapped by a volcano, whereupon crystallization resumes. This 'trick' of replenishment with fresh parental magma keeps the MgO content of the liquid from falling too rapidly and allows a much greater build-up of incompatible-element concentrations in the residual liquid than would be possible with closed-system fractional crystallization.

At first sight, these effects might seem to be of interest mostly to aficionados of the details of magma-chamber processes. But they imply a much-reduced role for chemical heterogeneity of the mantle, as well as for the effects of partial melting, because most of the incompatible-element variability is now ascribed to processes occurring in the magma chambers.

How plausible is this model? O'Neill and Jenner propose that a global assemblage of magma chambers exists, in which crystallization processes vary locally, but which as an ensemble conform to the O'Hara–Albarède model. Although this model is apparently quite successful in describing the global observations, the question of why these locally variable crystallization processes should average out to this idealized model remains a mystery.

The authors have tested their model by comparing predicted and measured partition coefficients of incompatible trace elements for crystals forming in magma chambers (the partition coefficient is the concentration of an element in a crystal divided by its concentration in the liquid). For the most part, the agreement is impressive, but barium and potassium are significant exceptions. These elements behave like highly incompatible elements in the basalts. In other words, their partition coefficients should be close to zero, which is actually the case in mantle minerals. But their experimentally determined partition coefficients in plagioclase (one of the main minerals that form in shallow magma chambers) are high enough to raise questions about the model.

Clearly, further experimental work is needed to resolve these issues. In the meantime, O'Neill and Jenner's paper indicates the need for a re-examination of the nature of a magmatic process that is volumetrically by far the most significant on Earth. An accurate assessment of the crystallization process is needed to infer the composition of the mantle from which the ocean-floor basalts are derived. ■

Albrecht W. Hofmann *is at the Max Planck Institute for Chemistry, 55020 Mainz, Germany, and at the Lamont-Doherty Earth Observatory, Columbia University, New York.*
e-mail: albrecht.hofmann@mpic.de

1. O'Neill, H. St C. & Jenner, F. E. *Nature* **491,** 698–704 (2012).
2. Jenner, F. E. & O'Neill, H. St C. *Geochem. Geophys. Geosyst.* **13,** Q02005 (2012).
3. Arevalo, R. Jr & McDonough, W. F. *Chem. Geol.* **271,** 70–85 (2010).
4. Sobolev, A. V. *et al. Science* **316,** 412–417 (2007).
5. O'Hara, M. J. *Nature* **266,** 503–507 (1977).
6. Albarède, F. *Nature* **318,** 356–358 (1985).

GENOMICS

Decoding our daily bread

The wheat genome is large and complex, and has defied complete sequencing. But the most comprehensive analysis so far of the plant's genes will support efforts to optimize the supply of this vital food crop. See Letter p.705

PETER LANGRIDGE

The bread wheat genome presents a significant challenge to researchers. At 17 gigabases, it is about six times the size of the human genome, and it is hexaploid, meaning that it contains six sets of chromosomes, which derive from three different genomes. So why bother to sequence such a difficult genome? Wheat is arguably the most important plant to humans. Bread wheat (*Triticum aestivum*) is the world's most widely grown crop, covering more than 200 million hectares of land[1] throughout temperate, Mediterranean-type and subtropical regions of both the Northern and Southern hemispheres. Although total production of wheat — 681 million tonnes in 2011 (ref. 1) — is slightly lower than that of maize (corn) and rice, wheat is the primary carbohydrate and protein source for the world's population[1]. For this reason, researchers around the world are tackling the challenge of the plant's genome. On page 705 of this issue, Brenchley *et al.*[2] present a detailed analysis and assembly of wheat gene sequences that will provide a key resource for crop scientists.

Systematic wheat breeding began around 100 years ago, but farmers' improvement of wheat strains by selective breeding can be traced back to the beginnings of agriculture almost 10,000 years ago[3]. The 'Green Revolution' of the 1960s — a series of advances in agricultural research, technology and infrastructure — triggered a drastic improvement in wheat yields. However, wheat production has struggled to meet global demand, and an increasingly variable and unstable climate is adding to the problems of wheat supply. It has been calculated that wheat production must increase by about 60% by 2050 to meet predicted demand[4]. This is a daunting challenge, but one that is taken seriously by the international community, as emphasized by the recent decision of the G20 group of countries to establish the international Wheat Initiative, designed to develop resources and capabilities to target wheat improvement[5]. A key objective of this initiative is to establish genomics resources so that new breeding technologies can be effectively and rapidly applied to wheat (Fig. 1).

The wheat genome is not only large, but also complex. It contains extensive stretches of repetitive non-coding DNA, which make sequence assembly difficult. Where there are genes, it is often hard to differentiate between the three constituent genomes, because each has a related set of genes. And the order of the genes has been partly shuffled on several of the chromosomes, adding to the complexity. There are three strategies that can be adopted to generate and assemble a full sequence for such a problematic genome. One approach is to make clone libraries that contain long stretches of DNA (more than 100,000 bases in each clone) derived from each wheat chromosome arm — there are 21 chromosomes (seven from each genome), giving 42 chromosome arms. The clones can be used to construct an overlapping series of DNA segments to produce a minimum tiling path for sequence assembly. This strategy is feasible because wheat is highly tolerant of chromosome changes and wheat lines are available in which each chromosome arm is present as a separate telosomic chromosome[6], which can be readily separated from the rest of the genome. Groups from around the world are working on this project as part of the International Wheat Genome Sequencing Consortium[7].

The second approach is to assemble the sequences of the three diploid genomes that are the progenitors of the wheat genome, which is broken down into A, B and D genomes. The progenitor species of the A and D genomes are known to be *Triticum urartu* and *Aegilops tauschii*, respectively, and sequence information is available. The progenitor of the B genome is believed to have been a close relative of *Aegilops speltoides*, and sequence information can be obtained from this species or derived from the tetraploid wheat species *Triticum durum*, which carries the A and B genomes[8].

'Shotgun sequencing' is the third approach. Unlike the large-clone process, which uses

OLIVER BRANDT/IMAGEBROKER/FLPA

Figure 1 | Vital grains. Brenchley and colleagues' detailed analysis[2] of the bread wheat genome will help scientists to identify breeding strategies to optimize yield from this crucial crop.

sizeable fragments to develop an approximate sequence 'map', the shotgun approach relies on sequence overlap in large numbers of much shorter sequences to assemble a genome. A public resource of shotgun sequences for each chromosomal arm of the wheat genome is close to completion[6]. However, although sequencing technologies are improving rapidly, such that shotgun sequencing of the entire hexaploid wheat genome is feasible, reliable assembly of these sequences from such a large genome is not yet possible.

The strategy of using large-insert clones to produce chromosome-arm-specific data is expected to yield the best-quality sequence assembly, but this will be a relatively slow process. However, the different approaches are not mutually exclusive and can be combined in a single effort, as Brenchley and colleagues have done. The authors' extensive sequencing led to the identification of between 94,000 and 96,000 genes. They compared these genes with sequence data from the progenitor genomes, and were able to assign around two-thirds of the genes to the A, B or D genomes. The approach was tested using sequences from individual chromosome arms — the researchers used shotgun sequencing of the isolated group 1 chromosomes (1A, 1B and 1D) to develop a set of sequences to 'train' the methods, then used them to assign sequences to specific genomes. The assignment of genes to the A, B or D genome is particularly valuable to wheat researchers because it allows them to differentiate genes and DNA markers from each of the three genomes, a difficult and time-consuming process.

Although Brenchley *et al.* have provided extensive sequence information, we are still a long way from having a complete wheat-genome assembly. However, the authors' data form a framework to which results from shotgun sequencing of other wheat varieties and from the chromosome-arm sequencing project can be added, and the reliability of the assembly will rise as more groups add their findings.

Brenchley and colleagues' sequence analysis also reveals the extent of wheat-genome flexibility. The researchers find that the formation of a hexaploid genome from three diploid progenitors has led to significant losses of members of many gene families, but also an expansion of other families, including those involved in plant metabolism and growth. These changes are likely to have been a key factor in the success of wheat in so many regions and climatic zones.

So will these findings give us clues to new strategies for wheat improvement? And can we harness the dynamism of the genome to generate varieties better able to cope with a variable environment? Wheat is grown largely in environments in which yield is being undermined by biotic and abiotic stresses. In fact, although wheat yields can exceed 12 tonnes per hectare (ref. 9), the global average is below 3 tonnes per hectare, and in non-irrigated environments the average is below 2 tonnes per hectare[1]. Heat and drought stress are foreshadowed as the major challenges for future wheat production, and plant breeding offers the best approach for responding to these pressures[10]. The current and future advances in understanding the wheat genome, and the genomes of other crop plants (Box 1), are likely to hold the key to developing breeding strategies that will optimize yields under variable conditions. ■

BOX 1

Insight into the barley genome

Anyone who likes a nip of whisky or a slurp of a malted milkshake will be pleased to know that researchers are also working towards preserving the world's supply of barley. On page 711 of this issue, the International Barley Genome Sequencing Consortium presents a genomic analysis for cultivated barley[11], *Hordeum vulgare* L., the world's fourth-most-abundant cereal crop.

The barley genome is less unwieldy than that of wheat, being only diploid and 5.1 gigabases in size. The authors used information from existing large-clone libraries to assemble a 4.98-gigabase physical map of the genome and then added data from whole-genome shotgun sequencing to this framework. Their analysis has identified more than 26,000 'high-confidence' genes (those for which similar genes have previously been identified in at least one other plant). It also reveals that substantial post-transcriptional processing, including high rates of alternative splicing, occurs to regulate gene expression in barley — as seen in wheat.

Barley is more stress tolerant than some other cereal plants, so it is able to grow comparatively well in harsh climates or on nutrient-poor soils. The genomic information may help to explain these resilience mechanisms, and facilitate the breeding of varieties that have optimal gene combinations for certain growth conditions. Plant breeders are also keen to identify genetic traits that might be exploited to further improve the dietary-fibre content of barley grain — which is already relatively high — because increased fibre intake may help to reduce the incidence of conditions such as type 2 diabetes and colorectal cancer. **Marian Turner**

Peter Langridge *is at the Australian Centre for Plant Functional Genomics, University of Adelaide, Urrbrae 5064, Australia.*
e-mail: peter.langridge@acpfg.com.au

1. FAOSTAT http://faostat3.fao.org (2012).
2. Brenchley, R. *et al. Nature* **491,** 705–710 (2012).

3. Zohary, D. & Hopf, M. *Domestication of Plants in the Old World* (Oxford Univ. Press, 2000).
4. Food and Agriculture Organization of the United Nations: Declaration of the World Summit on Food Security (Rome, 16–18 November 2009) (www.fao.org/wsfs/world-summit/en).
5. www.wheatinitiative.org (2012).
6. Dolezel, J., Kubaláková, M., Paux, E., Bartos, J. & Feuillet, C. *Chromosome Res.* **15,** 51–66 (2007).
7. www.wheatgenome.org (2012).
8. Feuillet, C., Langridge, P. & Waugh, R. *Trends Genet.* **24,** 24–32 (2008).
9. Lobell, D. B., Cassman, K. G. & Field, C. B. *Annu. Rev. Environ. Resour.* **34,** 179–204 (2009).
10. Tester, M. & Langridge, P. *Science* **327,** 818–822 (2010).
11. The International Barley Genome Sequencing Consortium *Nature* **491,** 711–716 (2012).

IMMUNOLOGY

Vitamins prime immunity

The finding that derivatives of vitamin B can bind to an antigen-presenting protein that stimulates specialized immune cells suggests a novel mechanism by which the immune system detects microbial infections. SEE ARTICLE P.717

WEI-JEN CHUA & TED H. HANSEN

In addition to their vital functions in development and metabolism, many vitamins have crucial roles in the immune system. The functions of two lipid-soluble vitamins, vitamin D (ref. 1) and vitamin A (ref. 2), in modulating immune responses are already known. But on page 717 of this issue, Kjer-Nielsen *et al.*[3] suggest a very different immune function for vitamins B2 (riboflavin) and B9 (folic acid), which are water-soluble vitamins. The authors provide evidence that molecules produced when bacteria metabolize certain B vitamins can activate a class of immune T cell called mucosa-associated invariant T (MAIT) cells. The proposal that MAIT cells detect infected cells through vitamin metabolites attached to the cells' surface is the first suggestion that vitamins can act as antigens (substances that activate T and B cells of the immune system), and boosts our understanding of this novel arm of the immune system.

T cells are major players in immunity, providing protection against infection. The most common ones are $CD4^+$ and $CD8^+$ T cells, which are found throughout the body and carry a broad repertoire of antigen receptors on their surface. $CD4^+$ and $CD8^+$ T-cell receptors bind to peptide antigens (fragments of proteins) that are 'displayed' on the surface of other cells by a cell-membrane protein belonging to the major histocompatibility complex (MHC) family. Development of these conventional $CD4^+$ and $CD8^+$ T cells depends on the presence of MHC molecules.

By contrast, MAIT cells are a type of unconventional T cell that are mostly found in the intestine, liver and lung[4] and that have only limited antigen-receptor diversity. MAIT-cell development depends on an MHC-related protein called MR1, which has been highly conserved over the course of mammalian evolution. Because the amino-acid sequence of MR1 is very similar to that of MHC molecules, it has been speculated that MR1 binds to specific antigens that lead to MAIT-cell activation. Indeed, genetic and biochemical studies suggest that MR1 presents antigen for MAIT-cell activation, but the chemical nature of the antigen was unknown[5].

Intriguingly, and uniquely among known T-cell populations, MAIT-cell survival also depends on the commensal microbiota — the non-pathogenic microorganisms that live on and in the body. Furthermore, MAIT cells are activated by the presence of cells infected with a diverse range of bacteria and yeast strains (although not viruses)[6,7]. Together, these findings hinted that MR1 is likely to bind microbial antigens, which are then presented to MAIT cells.

Kjer-Nielsen *et al.* have taken a key step in identifying exactly which antigens MR1

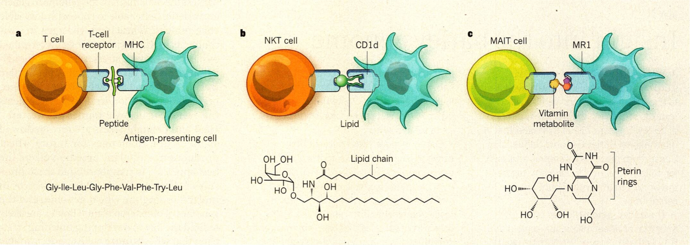


Figure 1 | Modes of antigen presentation. a, Conventional $CD4^+$ and $CD8^+$ T cells bind to antigens presented by MHC molecules on the surface of other cells. These antigens are typically short chains of amino acids (peptides) derived from proteins. The small balls extending from the peptide are amino-acid side chains that anchor the peptide in the MHC or are detected by the T-cell receptor. An example of a peptide derived from influenza virus that stimulates $CD8^+$ T cells is depicted below the cells. **b,** Another class of T cell, called NKT cells, recognizes antigens derived from lipid molecules that are presented by cells expressing a molecule called CD1d, which has deep grooves that can accommodate the lipid chains of the antigen. An example of a glycolipid that stimulates NKT cells is depicted. **c,** Kjer-Nielsen and colleagues[3] demonstrate that a third type of antigen-presenting molecule, called MR1, presents metabolites of B vitamins to T cells called MAIT cells. The authors present a crystal structure of the MR1 molecule bound to a derivative of vitamin B9, which shows that the antigen-binding groove accommodates the pterin-ring structures characteristic of B vitamins and their metabolites. An example of a vitamin-B2 metabolite that stimulates MAIT cells is shown.

presents by defining the crystal structure of MR1 bound by a metabolite of folic acid called 6-formyl pterin (6-FP). Their study stemmed from the serendipitous observation that using media containing folic acid enhanced the folding of denatured MR1 protein.

The authors' crystal structure shows that the MR1 antigen-binding groove specifically accommodates pterin rings, which are scaffold structures contained in some B vitamins and their metabolites. They also found that, although the 6-FP–MR1 complex does not activate MAIT cells *in vitro*, related riboflavin derivatives, when bound to MR1, do. This is the first demonstration that MR1 binds vitamin-B metabolites and that some of these metabolites can activate MAIT cells, and it therefore defines a new model of antigen presentation to immune cells. It was already known that MHC molecules present peptides to $CD4^+$ and $CD8^+$ T cells, and that another MHC-like protein called CD1d presents lipid molecules to a class of T cell called NKT cells; now we have evidence that MR1 presents vitamin-B metabolites to MAIT cells (Fig. 1).

However, exactly how this antigen-presentation process is linked to immunity to microbes is still not fully clear. Kjer-Nielsen and colleagues suggest that a mechanism by which MAIT cells detect and control infection is the display of vitamin-B metabolites on the surface of infected host cells. In support of this proposal, they cite the previous finding that the metabolic pathway that generates the antigenic molecules seems to be present only in microbes previously found to activate MAIT cells *in vitro*[6,7]. This correlative observation will need to be tested empirically to determine the importance of vitamin-B-metabolite presentation in controlling infection.

It will also be interesting to identify the cellular location and mechanism by which vitamin-B metabolites bind to MR1 proteins, and the role of infection in this process. In germ-free mice, which do not contain any commensal bacteria and therefore also lack MAIT cells, the addition of certain commensal bacterial strains allows MAIT cells to develop[7]. But, curiously, not all of these commensal strains have the molecular pathway that makes the metabolites studied by Kjer-Nielsen and colleagues, which implies that other MR1 ligands might be involved in microbial detection by MAIT cells. In addition, some cell-signalling molecules, such as interleukin-12 and interleukin-23, are known to activate MAIT cells[8,9], and this might mitigate the importance of activating signals derived from vitamin presentation by MR1.

Another pertinent question is that of the roles that MAIT cells have in the gut. Vitamins help to orchestrate the relationships between mammalian host immunity, commensal gut microbiota and pathogenic microorganisms[10]. For example, vitamin B9 and its derivatives can serve as coenzymes in essential metabolic pathways[11], and this vitamin is also required for the survival of a type of T cell called regulatory T cells[12]. Kjer-Nielsen and colleagues' findings suggest that interactions between the host and gut microbiota might also be influenced by MR1-dependent presentation of microbial antigens to MAIT cells.

A model that emerges from this idea is that, during early mammalian development, the colonization of the host by commensal bacteria[13] provides vitamin metabolites that act as ligands for MR1, allowing MAIT cells to develop in the thymus. These cells then migrate to other organs, in particular the lungs, liver and gut, where they help to guard against bacterial infections. Although further work is required, it is attractive to speculate that MAIT-cell-dependent protection against pathogens could be augmented by dietary provision of vitamins or by pterin-based therapies. Such augmentation might enhance immunity to microbes, or even help to treat immunodeficiencies. ■

Wei-Jen Chua *is at the US Food and Drug Administration, Bethesda, Maryland 20892, USA.* **Ted H. Hansen** *is at Washington University School of Medicine, St Louis, Missouri 63110, USA.*
e-mails: wei-jen.yankelevich@fda.hhs.gov; hansen@wustl.edu

1. Chun, R. F., Adams, J. S. & Hewison, M. *Expert Rev. Clin. Pharmacol.* **4,** 583–591 (2011).
2. Hall, J. A., Grainger, J. R., Spencer, S. P. & Belkaid, Y. *Immunity* **35,** 13–22 (2011).
3. Kjer-Nielsen, L. *et al. Nature* **491,** 717–723 (2012).
4. Treiner, E. *et al. Nature* **422,** 164–169 (2003).
5. Huang, S. *et al. Proc. Natl Acad. Sci. USA* **106,** 8290–8295 (2009).
6. Gold, M. C. *et al. PLoS Biol.* **8,** e1000407 (2010).
7. Le Bourhis, L. *et al. Nature Immunol.* **11,** 701–708 (2010).
8. Chua, W.-J. *et al. Infect. Immun.* **80,** 3256–3267 (2012).
9. Chiba, A. *Arthritis Rheum.* **64,** 153–161 (2012).
10. Nicholson, J. K. *et al. Science* **336,** 1262–1267 (2012).
11. Said, H. M. *Biochem. J.* **437,** 357–372 (2011).
12. Kunisawa, J., Hashimoto, E., Ishikawa, I. & Kiyono, H. *PLoS ONE* **7,** e32094 (2012).
13. Koenig, J. E. *et al. Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4578–4585 (2011).

QUANTUM PHYSICS

Strongly correlated transport

The field-effect transistor underlies microprocessor technology. A version of it has been demonstrated that tunes particle transport from an incoherent regime to a strongly correlated superfluid one. SEE LETTER P.736

LINCOLN D. CARR & MARK T. LUSK

The quantum transport of charge is a prominent feature in emerging models for future technologies, from microprocessors in computing to radical ideas for renewable-energy materials. In general, transport describes the motion of matter that has sufficient energy to overcome barriers, in contrast to quantum tunnelling, which occurs directly through barriers. Such transport becomes 'quantum' when the wave-like nature of particles is required to describe the process. There are different types of quantum transport, and devices that turn such flows on and off are called field-effect transistors (FETs)[1] and are a mainstay of microprocessor technology. In this issue, Stadler *et al.*[2] demonstrate a micrometre-scale FET that is able to capture the dynamics of quantum transport in a regime called strongly correlated quantum-coherent superfluidity. Their contribution can best be explained by tracking how different FETs can be used to study the dynamics of charge in a progression of transport regimes.

The basic concept of a FET is to use an electric field (gate) to control the transport of electrons or holes (notional particles formed by the absence of electrons) through a narrow channel between two charge reservoirs (source and drain) in a semiconducting material. Quantum mechanics must be used to predict the overall behaviour of the collection of charges, but the motion of each particle amounts to a (biased) random walk. The particles' dynamics is modelled by the semiclassical Boltzmann equation. However, under certain conditions, the wave-like nature of the charges becomes a major feature of transport properties. Then, different scattering paths can interfere like waves, building up constructive and destructive interference patterns, for instance, and the semi-classical Boltzmann equation is no longer correct.

On the edge of this transport regime, a blend of particle and wave-like dynamics is observed. Such partially coherent dynamics is currently being explored[3–5], for example to efficiently move charges in solar-cell materials and in the latest molecular-electronics devices (Fig. 1a,b). Deeper into the quantum-coherent transport regime, which is facilitated by very cold (near

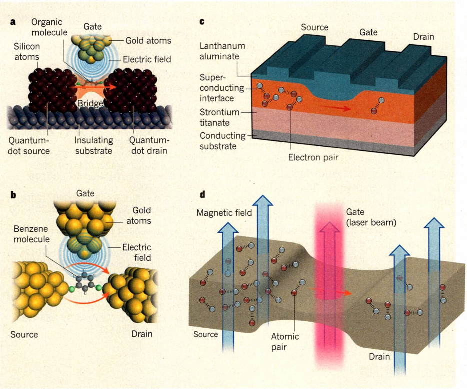


Figure 1 | Quantum transport in field-effect transistors (FETs). In a solid-state FET, an electrode called a gate, which is usually made of gold, generates an electric field that controls charge transport through a narrow channel between two reservoirs (source and drain). **a,** Partially coherent transport (orange) between quantum dots underlies a new generation of solar-cell materials[3,4], and quantum-dot FETs could be used to control and study the transport dynamics. The dots, which are made of assemblies of atoms (here silicon), are bridged by a small organic molecule and placed on an insulating substrate. The small green atoms connect the molecule to the reservoirs. **b,** A quantum-interference FET (or QuIET) uses the wave-like nature of the electron (not shown) in different paths (orange arrows) through a single benzene molecule to control charge flow from source to drain[5], both of which are made of gold. **c,** A FET based on a superconductor of electron pairs, formed at the interface of two oxides (lanthanum aluminate and strontium titanate), laid on a conducting substrate, at near-millikelvin temperatures[6] can be used to explore coherent transport and quantum-phase transitions. **d,** Stadler and colleagues[2] demonstrate a FET in which a laser beam serves as the device's gate and a magnetic field applied to the FET allows the interactions between ultracold atomic pairs to be tuned.

millikelvin) temperatures, the charges pair up into entities called Cooper pairs, their motion becomes weakly correlated (the motion of one pair depends on the motion of the others), and charge flows without resistance. This is the behaviour that underlies the properties of superconductors and their charge-neutral analogues, superfluids. A carefully crafted, cold version of the common FET has been developed for studying coherent-transport dynamics by forming such superconductors in complex oxide layers cooled in conventional dilution refrigerators[6] (Fig. 1c).

If the strength of the interactions between Cooper pairs is extreme, they become strongly correlated and the character of the system is fundamentally different. This is the regime of behaviour studied by Stadler and colleagues. The authors' FET is able to capture the dynamics of strongly correlated superfluids[7].

Instead of working with charges directly, their apparatus is designed to study the motion of neutral atoms that pair up to make diatomic molecules in place of the Cooper pairs of electrons in a superconductor. Their quantum-coherent FET comprises a source and drain that contain a gas of fermionic lithium atoms (Fig. 1d). Fermions have half-integer spin and are exclusive — that is, identical fermions refuse to be in the same quantum state. Stadler *et al.* trap their fermionic atoms in laser and magnetic fields, and at sub-microkelvin, ultracold temperatures — 10-million-fold colder than the cosmic microwave background radiation of outer space. Whereas FETs in microprocessors are typically tens of nanometres in size and have charge-carrier velocities approaching 10^7 centimetres per second, Stadler and colleagues' FET is on the scale of tens of micrometres and operates with transport velocities of between 1 millimetre and a few centimetres per second.

Significantly, the authors' apparatus allows them to tune the attractive atomic interactions over many orders of magnitude, so that the transport behaviour can be made to transition from the weakly correlated dynamics of the superconductors in the cold oxide FETs to the regime of strongly correlated superfluidity in ultracold fermionic atoms. Furthermore, the relatively large size of the authors' FET allows them to view what is actually happening in the transport channel, by means of high-resolution *in situ* optical imaging. The usual techniques for directly observing atoms, which involve atomic-force microscopy, cannot view such dynamics: we cannot zoom in on a solid-state FET in action.

Strongly correlated quantum behaviour has also been observed in high-energy nuclear-physics experiments[8–11] that create a state of matter known as a quark–gluon plasma, which is thought to have appeared shortly after the Big Bang. It has also been observed in ultracold quantum gases[12,13]. These two systems fall into the category of extreme quantum matter. Holographic duality[14–16] (a technique originating from string theory) infers that such strongly interacting quantum systems are mathematically equivalent to weakly curved gravity in one higher spatial dimension than our usual space-time continuum of four dimensions. This extra dimension has the physical effect of acting as a 'zoom' on a quantum system. Holographic duality predicts that these strongly correlated systems approach a perfect fluid, having a ratio of viscosity to entropy density far lower than that of ordinary fluids[7].

Stadler *et al.* thus offer a continuously tunable FET system that not only provides insight into quantum-coherent FET design, but also connects solid-state physics and extreme quantum matter in the new regime of quantum-coherent FET operation. ■

Lincoln D. Carr *and* **Mark T. Lusk** *are in the Department of Physics, Colorado School of Mines, Colorado 80401, USA.*
e-mail: lcarr@mines.edu

1. Cobbold, R. S. C. *Theory and Applications of Field-effect Transistors* (Wiley-Interscience, 1970).
2. Stadler, D., Krinner, S., Meineke, J., Brantut, J.-P. & Esslinger, T. *Nature* **491,** 736–739 (2012).
3. Lin, Z. *et al. ACS Nano* **6,** 4029–4038 (2012).
4. Engel, G. S. *et al. Nature* **446,** 782 (2007).
5. Stafford, C. A., Cardamone, D. M & Mazumdar, S. *Nanotechnology* **18,** 424014 (2007).
6. Caviglia, A. D. *et al. Nature* **456,** 624–627 (2008).
7. Adams, A., Carr, L. D., Schaefer, T., Steinberg, P. & Thomas, J. E. *N. J. Phys.* **14,** 115009 (2012).
8. BRAHMS Collaboration *Nucl. Phys. A* **757,** 1–27 (2005).
9. PHOBOS Collaboration *Nucl. Phys. A* **757,** 28–101 (2005).
10. STAR Collaboration *Nucl. Phys. A* **757,** 102–183 (2005).
11. PHENIX Collaboration *Nucl. Phys. A* **757,** 184–283 (2005).
12. O'Hara, K. M. *et al. Science* **298,** 2179–2182 (2002).
13. Cao, C. *et al. Science* **331,** 58–61 (2011).
14. Maldacena, J. M. *Adv. Theor. Math. Phys.* **2,** 231–252; also available at http://arxiv.org/abs/hep-th/9711200 (1998).
15. Gubser, S. S. *et al. Phys. Lett. B* **428,** 105–114 (1998).
16. Witten, E. *Adv. Theor. Math. Phys.* **2,** 253–291; also available at http://arxiv.org/abs/hep-th/9802150 (1998).

doi:10.1038/nature11574

Making sense of palaeoclimate sensitivity

PALAEOSENS Project Members*

Many palaeoclimate studies have quantified pre-anthropogenic climate change to calculate climate sensitivity (equilibrium temperature change in response to radiative forcing change), but a lack of consistent methodologies produces a wide range of estimates and hinders comparability of results. Here we present a stricter approach, to improve intercomparison of palaeoclimate sensitivity estimates in a manner compatible with equilibrium projections for future climate change. Over the past 65 million years, this reveals a climate sensitivity (in K W^{-1} m^2) of 0.3–1.9 or 0.6–1.3 at 95% or 68% probability, respectively. The latter implies a warming of 2.2–4.8 K per doubling of atmospheric CO_2, which agrees with IPCC estimates.

Characterizing the complex responses of climate to changes in the radiation budget requires the definition of climate sensitivity: this is the global equilibrium surface temperature response to changes in radiative forcing (an alteration to the balance of incoming and outgoing energy in the Earth–atmosphere system) caused by a doubling of atmospheric CO_2 concentrations. Despite progress in modelling and data acquisition, uncertainties remain regarding the exact value of climate sensitivity and its potential variability through time. The range of climate sensitivities in climate models used for Intergovernmental Panel for Climate Change Assessment Report 4 (IPCC-AR4) is 2.1–4.4 K per CO_2 doubling[1], or a warming of 0.6–1.2 K per W m^{-2} of forcing. Observational studies have not narrowed this range, and the upper limit is particularly difficult to estimate[2].

Large palaeoclimate changes can be used to estimate climate sensitivity on centennial to multi-millennial timescales, when estimates of both global mean temperature and radiative perturbations linked with slow components of the climate system (for example, carbon cycle, land ice) are available (Fig. 1). Here we evaluate published estimates of climate sensitivity from a variety of geological episodes, but find that intercomparison is hindered by differences in the definition of climate sensitivity between studies (Table 1). There is a clear need for consistent definition of which processes are included and excluded in the estimated sensitivity, like the need for strict taxonomy in biology. The definition must agree as closely as possible with that used in modelling studies of past and future climate, while remaining sufficiently pragmatic (operational) to be applicable to studies of different climate states in the geological past.

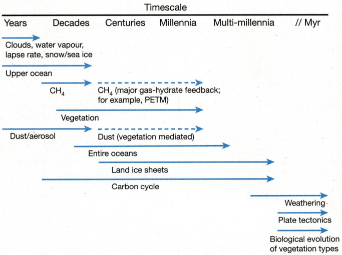

Timescale
Years
Decades
Centuries
Millennia
Multi-millennia
// Myr
Clouds, water vapour, lapse rate, snow/sea ice
Upper ocean
CH$_4$
CH$_4$ (major gas-hydrate feedback; for example, PETM)
Vegetation
Dust/aerosol
Dust (vegetation mediated)
Entire oceans
Land ice sheets
Carbon cycle
Weathering
Plate tectonics
Biological evolution of vegetation types

Figure 1 | Typical timescales of different feedbacks relevant to equilibrium climate sensitivity, as discussed in this work. Modified and extended from previous work[98]. Ocean timescales were extended to multi-millennial timescales[99].

Here we propose a consistent operational definition for palaeoclimate sensitivity and illustrate how a tighter definition narrows the range of reported estimates. Consistent intercomparison is crucial to detect systematic differences in sensitivity values—for example, due to changing continental configurations, different climate background states, and the types of radiative perturbations considered. These differences may then be evaluated in terms of additional controls on climate sensitivity, such as those arising from plate tectonics, weathering cycles, changes in ocean circulation, non-CO_2 greenhouse gases (GHGs), enhanced water-vapour and cloud feedbacks under warm climate states. Palaeoclimate data allow such investigations across geological episodes with very different climates, both warmer and colder than today. Clarifying the dependence of feedbacks, and therefore climate sensitivity, on the background climate state is a top priority, because it is central to the utility of past climate sensitivity estimates in assessing the credibility of future climate projections[1,3].

Quantifying climate sensitivity

'Equilibrium climate sensitivity' is classically defined as the simulated global mean surface air temperature increase (ΔT, in K) in response to a doubling of atmospheric CO_2, starting from pre-industrial conditions (which corresponds to a radiative perturbation, ΔR, of 3.7 W m^{-2}; refs 1, 3). We introduce the more general definition of the 'climate sensitivity parameter' as the mean surface temperature response to any radiative perturbation ($S = \Delta T / \Delta R$; where ΔT and ΔR are centennial to multi-millennial averages), which facilitates comparisons between studies from different time-slices in Earth history. For brevity, we refer to S as 'climate sensitivity'. In the definition of S, an initial perturbation ΔR_0 leads to a temperature response ΔT_0 following the Stefan–Boltzmann law, which is the temperature-dependent blackbody radiation response. This is often referred to as the Planck response[4], with a value S_0 of about 0.3 K W^{-1} m^2 for the present-day climate[5,6]. The radiative perturbation of the climate system is increased (weakened) by various positive (negative) feedback processes, which operate at a range of different timescales (Fig. 1). Because the net effect of positive feedbacks is found to be greater than that of negative feedbacks, the end result is an increased climate sensitivity relative to the Planck response[4].

---

*Lists of participants and their affiliations appear at the end of the paper.

Table 1 | Summary of key studies.

Label in Fig. 3	Source	Time window	Explicitly considered forcings	Temperature data used	S and 1σ bounds ($\mathrm{K\,W^{-1}\,m^2}$)	Notes
1	Ref. 2	LGM	Various	Various	0.81 ± 0.27 (data); $0.81^{+0.4}_{-0.27}$ (models)	LGM compilation based on ref. 15
2	Ref. 6	LGM	GHG (CO_2, CH_4, N_2O), LI, AE, VG	$\Delta T_{\mathrm{global}} = -5.8 \pm 1.4$ K; GLAMAP extrapolated with model[82]	$0.72^{+0.33}_{-0.23}$	Scaling factor (0.85) for smaller S at LGM compared to 2 × CO_2 (refs 12, 16)
3	Ref. 86	LGM	GHG (CO_2, CH_4), LI, AE, VG	CLIMAP and $\Delta T_{\mathrm{aa\&gld}}$	0.80 ± 0.14	Value after authors' suggested correction of CLIMAP temperatures
4	Ref. 79	LGM	GHG (CO_2, CH_4), LI, AE, VG	MARGO[81] SST based $\Delta T = -3.0^{+1.3}_{-0.7}$ K	$0.62^{+0.08}_{-0.12}$	Model-based global estimate
5	Ref. 76	GC	GHG (CO_2, CH_4)	ΔT_{trop}	1.1 ± 0.05	Author's linear regression case. Value based on single-site tropical SST, and representation of global changes will be more uncertain
6	Ref. 74	GC	GHG (CO_2, CH_4), LI, AE	ΔT_{aa} (with 1.5 × polar amplification)	0.88 ± 0.13	Author used a single value for polar amplification. If 2 × were used[52], then the central estimate is closer to 0.7
7	Ref. 52	GC	GHG (CO_2, CH_4, N_2O), LI	ΔT_{aa} (with 2 × polar amplification)	0.75 ± 0.13	Authors used a single value for polar amplification. If 1.5 × were used[74], then the central estimate becomes 1.0
8	Ref. 52	GC	GHG (CO_2, CH_4, N_2O)	ΔT_{aa} (with 2 × polar amplification)	1.5 ± 0.25	Authors used a single value for polar amplification. If 1.5 × were used[74], then the central estimate becomes 2.0
23–32	This work, based on ref. 6	GC (<800 kyr ago)	GHG (CO_2, CH_4, N_2O), LI, AE, VG	ΔT_{NH} = model-based deconvolution of benthic $\delta^{18}O$ (ref. 51), scaled to global ΔT using a NH polar amplification on land of 2.75 ± 0.25	0.66 ± 0.22 to 2.26 ± 0.78	This covers the range of $S_{[\mathrm{GHG,X}]}$ given in Table 2
9	Ref. 85	GC	GHG (CO_2, CH_4, N_2O), LI	ΔT_{aa} (with 2 × polar amplification) and 1.5 × ΔT_{ds}	0.75 ± 0.13	Authors used a single value for polar amplification. If 1.5 × were used[73], then the central estimate becomes 1.0
10	Ref. 39	GC	GHG (CO_2, CH_4, N_2O), LI, AE	36-record global SST synthesis along with $\Delta T_{\mathrm{aa\&gld}}$.	$0.85^{+0.25}_{-0.2}$	Polar amplification diagnosed, not imposed. Estimates made both in a spatially explicit sense and as direct global means
11	Ref. 39	GC	GHG (CO_2, CH_4, N_2O), LI	36-record global SST synthesis along with $\Delta T_{\mathrm{aa\&gld}}$.	1.05 ± 0.25	As above
12	Ref. 87	Early to Middle Pliocene (4.2–3.3 Myr ago)	CO_2, ESS	Using model-based ΔT for Middle and Early Pliocene of 2.4–2.9 °C and 4 °C. ΔCO_2 alkenone	1.92 ± 0.14 to 2.35 ± 0.18 (3.3 Myr ago); 2.60 ± 0.19 (4.2 Myr ago)	Forcing in ref. 44; temperature in ref. 87. Both derived in global sense from model experiments
13	Ref. 65	Miocene optimum to present day	Slow feedbacks	Deconvolution of benthic $\delta^{18}O$ (ref. 63)	0.78 ± 10%	$f = 0.71$, $\beta = 5.35$, $\gamma = 1.3$. Details in Supplementary Information
14	This work (compilation)	Eocene–Oligocene transition (~34 Myr ago)	CO_2. ESS (in the sense of ref. 44)	Model-based ΔT, with range of CO_2 values	$1.72^{+0.9}_{-0.54}$	Details in Supplementary Information
15	This work (compilation)	Late Eocene versus present	CO_2. ESS (in the sense of ref. 44)	Model-based ΔT, with range of CO_2 values	$1.82^{+0.26}_{-0.49}$	Details in Supplementary Information
16	Ref. 78	Middle Eocene Climatic Optimum (~40 Myr ago)	CO_2. Ice-free world. Event study (not affected by plate tectonics and evolution effects)	ΔT_{ds} (2 records) and ΔT_{mg} (7 records). ΔCO_2 from alkenones	0.95 ± 0.3	500 kyr timescale. $\Delta T_{\mathrm{ds}} = \Delta T_{\mathrm{mg}}$. Temperatures from subtropics to high latitudes; no tropical data. Hence biased to high-latitude sensitivity
17	Ref. 78	Mid to Late Eocene transition (41–35 Myr ago)	CO_2. Largely ice-free world. Event study (not affected by plate tectonics and evolution effects)	ΔT_{ds} (ref. 71) and ΔT_{mg}. ΔCO_2 = difference mid Eocene alkenone and late Eocene $\delta^{11}B$	0.95 ± 0.3	Multi-million-year timescale. Adding uncertainty of ±1 °C to ΔT would enhance 1σ limits to $\pm 0.45\ \mathrm{K\,W^{-1}\,m^2}$
18	Ref. 88	Early Eocene (~55–50 Myr ago)	CO_2. Ice-free world. (potential influences of plate tectonics and biological evolution not considered)	ΔT_{mg} (refs 89–91). ΔCO_2 based on modelling[91] marine organic carbon isotope fractionation[92] and soil nodules[93]	0.65 ± 0.25	Central value recalculated in ref. 94. Note ref. 89 underestimated tropical SST
19	This work (compilation)	PETM (~56 Myr ago)	CO_2. Ice-free world. Event study (not affected by plate tectonics and evolution effects)	ΔT_{ds} (>6 records) and ΔT_{mg} (>11 records; equatorial to polar). ΔCO_2 based on deep ocean carbonate chemistry[72, 95]	1.0–1.8	Details in Supplementary Information. Assumes all warming due to C input, and range of background CO_2 and C-injection scenarios. $\Delta T_{\mathrm{ds}} = \Delta T_{\mathrm{mg}}$. Total range of S is $0.7\text{–}2.2\ \mathrm{K\,W^{-1}\,m^2}$.

Table 1 | Continued

Label in Fig. 3	Source	Time window	Explicitly considered forcings	Temperature data used	S and 1σ bounds ($\mathrm{K\,W^{-1}\,m^2}$)	Notes
20	Ref. 96	Cretaceous and early Palaeogene	CO_2. Largely ice-free world. (potential influences of plate tectonics and biological evolution not considered)		1	Recalculated in ref. 94. No uncertainty range was reported, nor salient details for assessment. Figure 3b, c assumes ±25%
21	Ref. 94	Cretaceous and early Palaeogene	CO_2. Largely ice-free world. ESS in the sense of ref. 44	ΔT after refs 52, 71. ΔCO_2 based on ref. 60.	>0.8	No uncertainty range reported. This is a lower bound estimate only
22	Ref. 97	Phanerozoic	CO_2. Ice-free situation. (Potential influences of plate tectonics and biological evolution not considered).	ΔT_{mg}, ΔCO_2 based on GEOCARBSULF	0.8–1.08	Model-based with extensive uncertainty analysis

These studies have empirically determined S for the Pleistocene and some deep-time periods from comparison between data-derived time series for temperature and for radiative change. Comparison of results between studies is greatly hindered by the different 'versions' of S used, as related to different notions of which processes should be explicitly accounted for, and by the different approaches taken to approximate global mean surface temperature. All uncertainties are as originally reported, but shown here at the level equivalent to 1σ, estimated where necessary by dividing total range values by a factor of 2. All values for S are reported in $\mathrm{K\,W^{-1}\,m^2}$, where necessary after transformation using $3.7\,\mathrm{W\,m^{-2}}$ per doubling of CO_2, bearing in mind the caveats for this at high CO_2 concentrations as elaborated in the main text. GC, glacial cycles; LGM, Last Glacial Maximum; PETM, Palaeocene/Eocene thermal maximum; SST, sea surface temperature. See main text for details of forcings. Subscripts: aa, Antarctica; gld, Greenland; trop, tropical; ds, deep sea; global, global mean; mg, Mg/Ca; NH, Northern Hemisphere.

We emphasize that all feedbacks, and thus the calculated climate sensitivity, may depend in a—largely unknown—nonlinear manner on the state of the system before perturbation (the 'background climate state') and on the type of forcing[7–15]. The relationship of S with background climate state differs among climate models[12,16–18]. A suggestion of state dependence is also found in a data comparison (Table 2)[6], where climate sensitivity for the past 800,000 years (800 kyr) shows substantial fluctuations through time (Fig. 2). In contrast, its values for the Last Glacial Maximum (LGM) alone occupy only the lower half of that range (Fig. 2). That evaluation also suggests that the relationship of S with the general climate state may not be simple.

'Fast' versus 'slow' processes

Climate sensitivity depends on processes that operate on many different timescales, from seconds to millions of years, due to both direct response to external radiative forcing, and internal feedback processes (Fig. 1). Hence, the timescale over which climate sensitivity is considered is critical. An operationally pragmatic decision is needed to categorize a process as 'slow' or 'fast', depending on the timescale of interest, the resolution of the (palaeo-)records considered and the character of changes therein[19]. If a process results in temperature changes that reach steady state slower than the timescale of the underlying radiative perturbation, then it is considered 'slow'; if it is faster or coincident, then it is 'fast'. Present-day atmospheric GHG concentrations and the radiative perturbation due to anthropogenic emissions increase much faster than observed for any natural process within the Cenozoic era[20–22].

For the present, the relevant timescale for distinguishing between fast and slow processes can be taken as 100 yr (ref. 23). Ocean heat uptake plays out over multiple centuries. Combined with further 'slow' processes, it causes climate change over the next few decades to centuries to be dominated by the so-called 'transient climate response' to radiative changes that result from changing GHG concentrations and aerosols[5,19,24]. After about 100 yr, this transient climate response is thought to amount to roughly two-thirds of the equilibrium (see below) climate sensitivity[5,25]. Climate models account for the fast feedbacks from changes in water-vapour content, lapse rate, cloud cover, snow and sea-ice albedo[26], and the resulting response is often referred to as the 'fast-feedback' or 'Charney' sensitivity[23]. To approximate the 'equilibrium' value of that climate sensitivity, accounting for ocean heat uptake and further slow processes, models might be run over centuries with all the associated computational difficulties[27–30], or alternative approaches may be used that exploit the energy balance of the system for known forcing or extrapolation to equilibrium[31].

In palaeoclimate studies, an operational distinction has emerged to distinguish 'fast' and 'slow' processes relative to the timescales of temperature responses measured in palaeodata, where 'fast' is taken to apply to processes up to centennial scales, and 'slow' to processes with timescales close to millennial or longer. Thus, changes in natural GHG concentrations are dominated by 'slow' feedbacks related to global biogeochemical cycles (Fig. 1). Similarly slow are the radiative influences of albedo feedbacks that are dominated by centennial-scale or longer changes in global vegetation cover and global ice area/volume (continental ice sheets) (Fig. 1).

Table 2 | Common permutations of *S* that may be encountered in palaeostudies

Label in Fig. 3	S definition	Explicitly considered radiative perturbation	Period in which it is practical to use the definition	$S \pm 1\sigma$ for 800 kyr ($\mathrm{K\,W^{-1}\,m^2}$)	$S \pm 1\sigma$ for LGM ($\mathrm{K\,W^{-1}\,m^2}$)	S for Pliocene ($\mathrm{K\,W^{-1}\,m^2}$)
23	$S_{[\mathrm{CO2}]}$	$\Delta R_{[\mathrm{CO2}]}$	All (especially pre-35 Myr ago when LI ≈ 0)	3.08 ± 0.96	2.63 ± 0.57	1.2
24	$S_{[\mathrm{CO2,\,LI}]}$	$\Delta R_{[\mathrm{CO2,\,LI}]}$	<35 Myr ago	1.07 ± 0.40	0.95 ± 0.22	0.97
25	$S_{[\mathrm{CO2,\,LI,\,VG}]}$	$\Delta R_{[\mathrm{CO2,\,LI,\,VG}]}$	<35 Myr ago	0.86 ± 0.27	0.80 ± 0.19	0.82
26	$S_{[\mathrm{CO2,\,LI,\,AE}]}$	$\Delta R_{[\mathrm{CO2,\,LI,\,AE}]}$	<35 Myr ago, but mainly <800 kyr ago	0.90 ± 0.42	0.72 ± 0.18	
27	$S_{[\mathrm{CO2,\,LI,\,AE,\,VG}]}$	$\Delta R_{[\mathrm{CO2,\,LI,\,AE,\,VG}]}$	<35 Myr ago, but mainly <800 kyr ago	0.75 ± 0.29	0.63 ± 0.15	
28	$S_{[\mathrm{GHG}]}$	$\Delta R_{[\mathrm{GHG}]}$	<800 kyr ago	2.32 ± 0.76	1.97 ± 0.41	
29	$S_{[\mathrm{GHG,\,LI}]}$	$\Delta R_{[\mathrm{GHG,\,LI}]}$	<800 kyr ago	0.96 ± 0.36	0.85 ± 0.19	
30	$S_{[\mathrm{GHG,\,LI,\,VG}]}$	$\Delta R_{[\mathrm{GHG,\,LI,\,VG}]}$	<800 kyr ago	0.78 ± 0.23	0.73 ± 0.16	
31	$S_{[\mathrm{GHG,\,LI,\,AE}]}$	$\Delta R_{[\mathrm{GHG,\,LI,\,AE}]}$	<800 kyr ago	0.82 ± 0.36	0.66 ± 0.16	
32	$S_{[\mathrm{GHG,\,LI,\,AE,\,VG}]}$	$\Delta R_{[\mathrm{GHG,\,LI,\,AE,\,VG}]}$	<800 kyr ago	0.68 ± 0.24	0.58 ± 0.14	

S (second column) is presented with a subscript that identifies the explicitly considered radiative perturbations ΔR (third column, same subscripts as for S); all other processes are implicitly resolved as feedbacks within S. The period in which the various definitions of S are practical is determined by the availability of data for the explicitly considered processes. Subscript CO2 indicates the radiative impact of atmospheric CO_2 concentration changes; LI represents the radiative impact of global land ice-volume changes; VG stands for the radiative impact of global vegetation cover changes; AE indicates the radiative impact of aerosol changes; GHG stands for the impact of changes in all non-water natural greenhouse gases (notably CO_2, CH_4 and N_2O). Columns 5 and 6 give calculated values for all suggested permutations of S for the past 800 kyr or the LGM, respectively, based on a previous data compilation[6]. Mean values of all $S_{[X]}$ for the LGM are about 13% smaller than for the whole 800 kyr, but lie well within the given uncertainties. This offset illustrates the state-dependence of S (see Supplementary Information). Column 7 gives examples for the Pliocene[13,44]; Fig. 3b, c assumes ±25% uncertainty in these. In these values the effects of orographic changes have been taken into account (see Supplementary Information section B2).

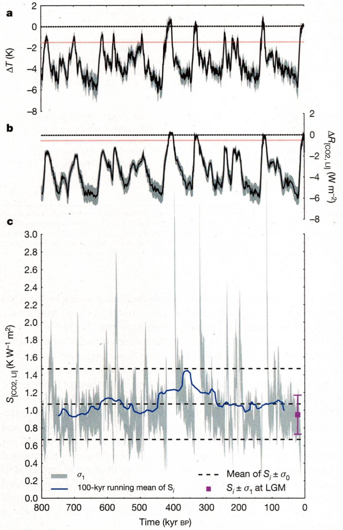


Figure 2 | Illustration of variability of climate sensitivity using a calculation of $S_{[CO2,LI]}$, as defined in this work, for the past 800 kyr. **a**, Changes in global temperature. **b**, Changes in radiative forcing due to changes in CO_2 and surface albedo due to land ice. **c**, Calculated $S_{[CO2,LI]}$, which is only considered robust and calculated when $\Delta T < -1.5$ K and $\Delta R_{[CO2,LI]} < -0.5$ W m^{-2}, as indicated by the dotted red lines in **a** and **b**. In **c**, mean of $S_i \pm \sigma_0$ (dashed black lines indicate σ_0, the uncertainty of averaging) and 100-kyr running mean (blue line) are shown. Magenta marker in **c** denotes $S_i \pm \sigma_1$ for the LGM only (23–19 kyr ago) (σ_1 is the square root of the sum of squares of individual uncertainties connected with different processes contributing to S_i). The grey areas in **a**–**c** denote σ_1 (standard deviation) uncertainties of S_i for single points in time (points themselves are omitted for clarity). Details of data and the definition of the calculated uncertainties presented in this figure are available in Supplementary Information. In **a** and **b**, the dashed black lines indicate the preindustrial reference case ($\Delta T = 0$ K, $\Delta R_{[CO2,LI]} = 0$ W m^{-2}).

Other processes clearly have both fast and slow components. For example, palaeorecords of atmospheric dust deposition imply important aerosol variations on decadal to astronomical (orbital) timescales[32–36], reflecting both slow controlling processes related to ice-volume and land-surface changes, and fast processes related to changes in atmospheric circulation. A further complication arises from the lack of adequate global atmospheric dust data for any geological episode except the LGM[37,38], even though that is essential because the spatial distribution of dust in the atmosphere tends to be inhomogeneous and because temporal variations in some locations tend to take place over several orders of magnitude[32–36]. Moreover, palaeoclimate models generally struggle to account for aerosols, with experiments neither prescribing nor implicitly resolving aerosol influences. So far, understanding of aerosol/dust feedbacks remains weak and in need of improvements in both data coverage and process modelling, especially because dust forcing may account for some 20% of the glacial–interglacial change in the radiative budget[6,39].

So for comparison of results between studies, it is most effective to consider only the classical 'Charney' water-vapour, cloud, lapse rate, and snow and sea-ice feedbacks[23] as 'fast', and all other feedbacks as 'slow'. In addition, results from palaeoclimate sensitivity studies generally do not address the transient climate response that dominates present-day changes, but capture a more complete longer-term system response comparable with equilibrium climate sensitivity in climate models.

Forcing and slow feedbacks

The external drivers of past natural climate changes mainly resulted from changes in solar luminosity over time[40], from temporal and spatial variations in insolation due to changes in astronomical parameters[41–43], from changes in continental configurations[14,44], and from geological processes that directly affect the carbon cycle (for example, volcanic outgassing). However, the complete Earth system response to such forcings as recorded by palaeodata cannot be immediately deduced from the (equilibrium) 'fast feedback' sensitivity of climate models, because of the inclusion of slow feedback contributions. When estimating climate sensitivity from palaeodata, agreement is therefore needed about which of the slower feedback processes are viewed as feedbacks (implicitly accounted for in S), and which are best considered as radiative forcings (explicitly accounted for in ΔR).

We employ an operational distinction[31,45] in which a process is considered as a radiative forcing if its radiative influence is not changing with temperature on the timescale considered, and as a feedback if its impact on the radiation balance is affected by temperature changes on that timescale. For example, the radiative impacts of GHG changes over the past 800 kyr may be derived from concentration measurements of CO_2, CH_4 and N_2O in ice cores[46–48], and the radiative impacts of land-ice albedo changes may be calculated from continental ice-sheet estimates, mainly based on sea-level records[49–51]. Thus, the impacts of these slow feedbacks can be explicitly accounted for before climate sensitivity is calculated. This leaves only fast feedbacks to be considered implicitly in the calculated climate sensitivity, which so approximates the (equilibrium) 'Charney' sensitivity from modelling studies[6,39,52].

Operational challenges

All palaeoclimate sensitivity studies are affected by limitations of data availability. Below we discuss such limitations to reconstructions of forcings and feedbacks, and of global surface temperature responses. First, however, we re-iterate a critical caveat, namely that the climate response depends to some degree on the type of forcing (for example, shortwave versus longwave, surface versus top-of-atmosphere, and local versus global). The various radiative forcings with similar absolute magnitudes have different spatial distributions and physics, so that the concept of global mean radiative forcing is a simplification that introduces some (difficult to quantify) uncertainty.

Astronomical (orbital) forcing is a key driver of climate change. In global annual mean calculations of radiative change, astronomical forcing is very small and often ignored[39,52]. Although this obscures its importance, mainly concerning seasonal changes in the spatial distribution of insolation over the planet[41,42,53–55], we propose that the contribution of the astronomical forcing to ΔR may be neglected initially. When other components of the system respond to the seasonal aspects of forcing, such as Quaternary ice-sheet variations, these may be accounted for as forcings themselves.

GHG concentrations from ice cores are not available for times before 800 kyr ago, when CO_2 levels instead have to be estimated from indirect methods. These employ physico-chemical or biological processes that depend on CO_2 concentrations, such as the abundance of stomata on fossil leaves[56], fractionation of stable carbon isotopes by marine phytoplankton[57], boron speciation and isotopic fractionation in sea water as a function of pH and preserved in biogenic calcite[58], and the stability fields of minerals precipitated from waters in contact with the atmosphere[59].

Considerable uncertainties remain in such reconstructions, but improvements are continually made to the methods, their temporal coverage and their mutual consistency[60]. Recent work has synthesized a high-resolution CO_2 record for the past 20 million years (Myr; ref. 61), but new data and updated syntheses remain essential, particularly for warmer climate states. Also, proxies are needed for reconstruction of CH_4 and N_2O concentrations in periods pre-dating the ice-core records[62].

Regarding the assessment of land-ice albedo changes, good methods exist for the generation of continuous centennial- to millennial-scale sea-level (ice-volume) records over the past 500 kyr (refs 49–51), but such detailed information remains scarce for older periods. A model-based deconvolution of deep-sea stable oxygen isotope records into their ice-volume and deep-sea temperature components[51] was recently extended to 35 Myr ago[63], but urgently requires independent validation, especially to address uncertainties about the volume-to-area relationships that would be different for incipient ice sheets than for mature ice sheets[64,65]. Before 35 Myr ago, there is thought to have been (virtually) no significant land-ice volume[66], but this does not exclude the potential existence of major semi-permanent snow/ice-fields[67,68], and there remain questions whether these would constitute 'fast' (snow) or 'slow' (land-ice) feedbacks. The contribution of the sea-ice albedo feedback also remains uncertain, with little quantitative information beyond the LGM.

Similar examples of uncertainties and limited data availability could be listed for all feedbacks. However, a 'deep-time' (before 1 Myr ago) geological perspective must be maintained because it offers access to the nearest natural approximations of the current rate and magnitude of GHG emissions[69,70], and because only ancient records provide insight into climate states globally warmer than the present. Given that no past perturbation will ever present a perfect analogue for the continuing anthropogenic perturbation, it may be more useful to consider past warm climate states as test-beds for evaluating processes and responses, and for challenging/validating model simulations of those past climate states. Such data–model comparisons will drive model skill and understanding of processes, improving confidence in future multi-century projections. For such an approach, palaeostudies may minimize the impacts of very long-term influences on climate sensitivity (for example, due to changes in orography, or biological evolution of vegetation) through a focus on highly resolved documentation of specific perturbations that are superimposed upon different long-term background climate states. An example is the pronounced transient global warming and carbon-cycle perturbation during the Palaeocene/Eocene thermal maximum (PETM) anomaly[71,72], which punctuated an already warm climate state[73]. Note that deep-time case studies need to consider one further complication, namely that the radiative forcing per CO_2 doubling may be about $3.7\,W\,m^{-2}$ when starting from pre-industrial concentrations, but increases at higher CO_2 levels[11]. Data-led studies may help with a first-order documentation of this dependence. Calculation of S from CO_2 and temperature measurements using a constant $3.7\,W\,m^{-2}$ per CO_2 doubling would (knowingly) overestimate S for high-CO_2 episodes. The difference with other, identically defined, S values for different climate background states may then be used to assess any deviation from $3.7\,W\,m^{-2}$ per CO_2 doubling.

Regarding the reconstruction of past global surface temperature responses (that is, ΔT in equation (1) below), again much remains to be improved. Most work to date (see Table 1) relies on one or more of the following: polar temperature variations from Antarctic ice cores (since 800 kyr ago) with a multiplicative correction for 'polar amplification' (usually estimated at 1.5–2.0; refs 74, 75); deep-sea temperature variations from marine sediment-core data with a correction for the ratio between global surface temperature and deep-sea temperature changes (often estimated at 1.5); single-site sea surface temperature (SST) records from marine sediment cores; or compilations of SST data of varying geographic coverage from marine sediment cores[6,39,52,76–78]. So far, few studies have included terrestrial temperature proxy records other than those from ice cores[79], yet better control on land-surface data is crucial because of seasonal and land-sea contrasts. Continued development is needed of independently validated (multi-proxy) and spatially representative (global) data sets of high temporal resolution relative to the climate perturbations studied.

Uncertainties in individual reconstructions of temperature change may in exceptional cases be reported to ±0.5 K, but more comprehensive uncertainty assessments normally find them to be larger[80,81]. Compilation of such records to determine changes in global mean surface temperature involves the propagation of further assumptions/uncertainties, for example due to interpolation from limited spatial coverage, and the end result is unlikely to be constrained within narrower limits than ±1 K even for well-studied intervals. Finally, comparisons between independent reconstructions for the same episode reveal 'hidden' uncertainties due to differences between each study's methodological choices, uncertainty determination, and data-quality criteria, which are hard to quantify and often poorly elucidated. Take the LGM for example, which for temperature is among the best-studied intervals. The MARGO compilation[81] inferred a global SST reduction of $-1.9 \pm 1.8\,K$ relative to the present. Another spatially explicit study[79] used that range to infer a global mean surface air temperature anomaly of $-3^{+1.3}_{-0.7}K$. The latter contrasts with a previous estimate of $-5.8 \pm 1.4\,K$ (ref. 82), which is consistent with tropical (30° S to 30° N) SST anomalies of $-2.7 \pm 1.4\,K$ (ref. 83). However, that tropical range itself is also contested; the MARGO[81] study suggested such cooling in the Atlantic Ocean, but less in the tropics of the Indian and Pacific Oceans (giving a global tropical cooling of only $-1.7 \pm 1.0\,K$). Clearly, even a well-studied interval gives rise to a range of estimates for temperature, and therefore for climate sensitivity.

It is evident that progress in quantifying palaeoclimate sensitivity will not only rely on a common concept and terminology that allows like-for-like comparisons (see below); it will also rely on an objective, transparent and hence reproducible discussion in each study of the assumptions and uncertainties that affect the values determined for change in both temperature and radiative forcing.

A way forward

Here we propose a new terminology to help palaeoclimate sensitivity studies adopt common concepts and approaches, and thus improve the potential for like-for-like comparisons between studies. First we outline how our concept of 'equilibrium' S for palaeo-studies relates to 'equilibrium' S for modern studies. Then, we present a notation system that is primarily of value to data-based palaeo studies to clarify which slow feedbacks are explicitly accounted for. We finish with an application of the new framework, calculating climate sensitivity from a representative selection of palaeoclimate sensitivity estimates over the past 65 Myr, with a fair balance of climates warmer than the present to those colder than the present.

When the ΔT response to an applied GHG radiative forcing ΔR is small relative to 'pre-perturbation' reference temperature, the 'equilibrium' climate sensitivity S^a (where a indicates *actuo*, for present-day) takes the form (see, for example, refs 4, 84):

$$S^a = \frac{\Delta T}{\Delta R} = \frac{-1}{\lambda_P + \sum_{i=1}^{N} \lambda_i^f} \tag{1}$$

Here λ_P is the Planck feedback parameter ($-3.2\,W\,m^{-2}\,K^{-1}$) and λ^f_i (in $W\,m^{-2}\,K^{-1}$) represents the feedback parameters of any number (N) of fast (f) feedbacks. We define feedback parameters in the form $\lambda^f_i = \Delta R_i/\Delta T$. S^a is the 'Charney' sensitivity calculated by most climate models in '$2 \times CO_2$' equilibrium simulations, with a range of $0.6–1.2\,K\,W^{-1}\,m^2$ in IPCC-AR4. However, the Earth system in reality responds to a perturbation according to an equilibrium climate sensitivity parameter S^p (where p indicates *palaeo*), but the timescales to reach this equilibrium are long, so that the forcing normally changes before equilibrium is reached. To obtain S^a from palaeoclimate sensitivity S^p, a correction is therefore needed for the slow feedback influences. Using λ^s_j to represent any number (M) of slow (s) feedbacks, we derive the general expression (see Supplementary Information):

$$S^a = S^p\left(1+\frac{\sum_{j=1}^{M}\lambda_j^{\mathrm{s}}}{\lambda_{\mathrm{P}}+\sum_{i=1}^{N}\lambda_i^{\mathrm{f}}}\right) \qquad (2)$$

This approach is contingent on the above-mentioned caveats of state-dependence, linearization (small ΔT), changes in slow feedbacks, and transient effects, where the last is relevant only in records of exceptionally high temporal resolution. Knowledge of slow (λ^{s}) and fast (λ^{f}) feedbacks can be combined into a factor $F = \lambda^{\mathrm{s}}/(\lambda^{\mathrm{f}} + \lambda^{\mathrm{s}})$ that may then be used to back-calculate fast feedbacks out of palaeoclimate sensitivity S^p.

A recent study[44] defined the term 'Earth system sensitivity' (ESS) to represent the long-term climate response of Earth's climate system to a given CO_2 forcing, including both fast and slow processes. In our notation, ESS = $\Delta R_{2\times CO2}\, S^p$, where $\Delta R_{2\times CO2}$ is the forcing due to a CO_2-doubling (3.7 W m^{-2}).

Here we introduce a more explicit notation regarding what was (not) included in the climate sensitivity diagnosis. It is the 'specific climate sensitivity' $S_{[A,B\ldots]}$, expressed in K W^{-1} m^2, where slow feedback processes A, B, and so on, are explicitly accounted for (that is, included in the forcing term, $\Delta R_{[A,B\ldots]}$). We use 'LI' to denote albedo forcing due to land-ice volume/area changes, 'VG' for vegetation-albedo forcing, 'AE' for aerosol forcing and 'CO2' for atmospheric CO_2 forcing (see also Table 1). This approach requires from the outset that a comprehensive view is taken of the various causes of change in the radiative balance.

The most practical version of S to be estimated from palaeodata is $S_{[CO2,LI]}$, because $S_{[CO2,LI]} = S_{[CO2]}$ during times (pre-35 Myr ago) without ice volume, and because global vegetation cover changes, atmospheric dust fluctuations, and both CH_4 and N_2O fluctuations (the two important non-CO_2 GHGs) generally remain poorly constrained by proxy data. Common reporting of $S_{[CO2,LI]}$ would bring results closer in line with the model-based concept of 'equilibrium' fast-feedback sensitivity. The above-mentioned issues with aerosol influences mean that it would currently be best for estimates from palaeodata to report both $S_{[CO2,LI]}$ and $S_{[CO2,LI,AE]}$.

Table 2 lists example estimates for S following the main potential permutations of the definition of S in our approach (for detailed breakdowns, see Supplementary Information). The first example uses records of palaeodata since 800 kyr ago. The second example uses the same input data series[6], but focuses only on the LGM; the contrast between examples one and two thus highlights state-dependence. The third example lists estimates for $S_{[CO2]}$, $S_{[CO2,LI]}$ and $S_{[CO2,LI,VG]}$ from a more model-based assessment for the mid-Pliocene (~3–3.3 Myr ago)[13], with $\Delta T = 3.3$ K relative to the present and $\Delta R_{CO2} = 1.9$ W m^{-2} due to CO_2 increase from 280 to 400 parts per million by volume (p.p.m.v.; ref. 44). The broad range of S values found within each example illustrates that comparison across different definitions unrealistically widens the range of values reported, notably towards high values, because omission of 'forcing' due to the action of any slow feedbacks will cause overestimation of S (see also Fig. 3).

For a first-order estimate of the range of S from palaeodata that approximates compatibility with the centennial timescale 'equilibrium'

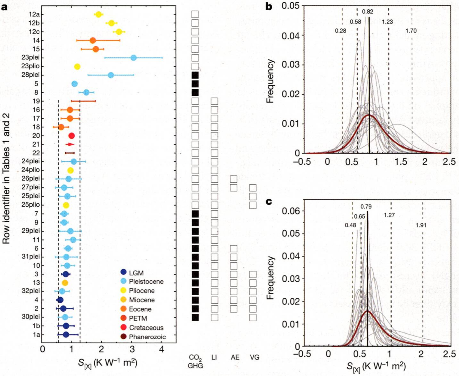


Figure 3 | Evaluation of results from Tables 1 and 2. *y*-Axis labels refer to numbered rows in these Tables. **a**, Data summary by table row. **b**, Probability assessment using normal distributions (shifted where relevant). **c**, Probability assessment using lognormal distributions. $S_{[X]}$ refers to the climate sensitivity as defined in detail by the subscript X in Tables 1 and 2. For **b** and **c**, we assume a relative uncertainty of 25% for entries that lacked uncertainty estimates in the source studies. In **a**, rows from Table 2 are identified with either 'plei' or 'plio' to distinguish between the past 800 kyr and the Pliocene entries, respectively. The colour coding refers to broad geological intervals, as shown in the key. Boxes at right indicate which conditions were explicitly accounted for; that is, as 'forcings' (in the CO_2/GHG column, filled squares indicate GHG and open squares CO_2). Circles (data points in **a**) show central values where reported, error bars represent uncertainties as outlined in the Tables, at the 1σ equivalent level. Arrow (case 21) indicates a value reported only as >0.8 K W^{-1} m^2. Black dashed lines in **a** show 68% probability limits for all estimates that account for at least 'CO2' and 'LI', based on thick dashed lines in **b** and **c**, taking whichever 68% value offers the widest (more conservatively estimated) margin. In **b** and **c**, the solid black line indicates the mode value (maximum), and the thin dashed lines the 95% probability limits. All distributions in **b** and **c** are given as individual normalized frequencies (grey lines), and as mean normalized frequencies (red line).

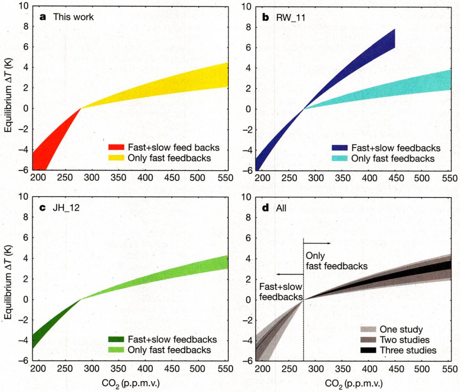


Figure 4 | Equilibrium response of the global temperature as a function of CO_2 concentrations, based on three different approaches. **a**, This work, using data from the late Pleistocene of the past 800 kyr (ref. 6). **b**, Using data of the past 20 Myr (RW_11; ref. 61). **c**, Based on JH_12 (ref. 85) using similar data of the past 800 kyr as in **a**. **d**, Combination of all three approaches. Plotted areas include uncertainty estimates of one standard deviation. Because this work and JH_12 developed their approach only on Pleistocene data (climate being mainly colder than today), extrapolation of the impact of slow feedbacks to $2 \times CO_2$ is not meaningful (we show only extrapolation with fast feedbacks). RW_11 in contrast also includes warmer climates with CO_2 up to 450 p.p.m.v., so that the applicable range with slow feedbacks extends to 450 p.p.m.v. For future climate with $2 \times CO_2$ and a short time horizon (<100 yr), only fast feedbacks are of interest (see **d**). Approaches partly disagree because of different assumptions. Uncertainties in this work (**a**) are estimated to be larger than they were in RW_11 (**b**) and JH_12 (**c**). For details of the equations and values used, see Supplementary Information.

values of the IPCC-AR4[1], values need to be used that account for 'CO_2' or 'GHG' as well as 'LI', and preferably also 'AE' and/or 'VG' (Tables 1, 2; Fig. 3). Such an assessment, excluding the case of row 21 in Table 1, yields a likely[1] (68%) probability range of 0.6–1.3 $K\,W^{-1}\,m^2$, and a 95% range of 0.3–1.9 $K\,W^{-1}\,m^2$ (Fig. 3). These represent the widest margins out of two assessments, using either normal distributions with shifts when relevant (Fig. 3a), or lognormal distributions that inherently allow asymmetry[2] (Fig. 3b). These assessments include uncertainties as outlined in the source studies, as well as any unaccounted-for dependence on different background climate states, but exclude potential additional uncertainties highlighted in this study. Inclusion of ESS values (approximated by $S_{[CO_2]}$) would extend the upper limit beyond 3 $K\,W^{-1}\,m^2$ (Fig. 3a). Future work following a strict framework for reporting and comparison of palaeodata may refine the observed asymmetry.

Finally, following our conceptual framework, we can make a projection of equilibrium temperature change over a range of CO_2 concentrations while considering either slow and fast (or only fast) feedbacks (Fig. 4; see Supplementary Information for details). Including the known uncertainties associated with palaeoclimate sensitivity calculations, and comparing with two previous approaches[61,85], we find overlap in the 68% probability envelopes that implies equilibrium warming of 3.1–3.7 K for $2 \times CO_2$ (Fig. 4), equivalent to a fast feedback (Charney) climate sensitivity between 0.8 and 1.0 $K\,W^{-1}\,m^2$. For longer, multi-centennial projections, some of the slow feedbacks (namely vegetation-albedo and aerosol feedbacks) may need further consideration. However, their impact is difficult to estimate from palaeodata, because uncertainties are large, and because responses during climates colder than present may differ from responses during future warming.

We have employed a new framework of definitions for palaeoclimate sensitivity. This reveals how a broad selection of previously published estimates for the past 65 Myr agrees on a best general estimate of 0.6–1.3 $K\,W^{-1}\,m^2$, which agrees with IPCC-AR4 estimates for equilibrium climate sensitivity[1]. Higher estimates than ours may suggest different climate sensitivities during particular periods, but a considerable portion of the higher values may simply reflect differences in the definitions of palaeoclimate sensitivity that were used.

Received 18 April; accepted 11 September 2012.

1. Solomon, S. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
2. Knutti, R. & Hegerl, G. C. The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nature Geosci.* **1,** 735–743 (2008).
   Presents a synthesis of equilibrium climate sensitivity estimates and discusses challenges for constraining its upper limit.
3. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
4. Roe, G. H. Feedbacks, timescales and seeing red. *Annu. Rev. Earth Planet. Sci.* **37,** 93–115 (2009).
5. Dufresne, J.-L. & Bony, S. An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. *J. Clim.* **21,** 5135–5144 (2008).
   Presents a compilation of results of 12 GCMs used in IPCC-AR4, on the contribution of different fast feedbacks to both equilibrium and transient temperature change.
6. Köhler, P. *et al.* What caused Earth's temperature variations during the last 800,000 years? Data-based evidences on radiative forcing and constraints on climate sensitivity. *Quat. Sci. Rev.* **29,** 129–145 (2010).
   Presents a data compilation on radiative forcing over the past 800 kyr, which forms the backbone of our late Pleistocene examples in Table 2 and in Supplementary Information.
7. Roe, G. H. & Baker, M. B. Why is climate sensitivity so unpredictable? *Science* **318,** 629–632 (2007).

8. Baker, M. B. & Roe, G. H. The shape of things to come: why is climate change so predictable? *J. Clim.* **22,** 4574–4589 (2009).
9. Hannart, A., Dufresne, J.-L. & Naveau, P. Why climate sensitivity may not be so unpredictable. *Geophys. Res. Lett.* **36,** L16707 (2009).
10. Zaliapin, I. & Ghil, M. Another look at climate sensitivity. *Nonlinear Process. Geophys.* **17,** 113–122 (2010).
11. Colman, R. & McAvaney, B. Climate feedbacks under a very broad range of forcing. *Geophys. Res. Lett.* **36,** L01702 (2009).
12. Hargreaves, J. C., Abe-Ouchi, A. & Annan, J. D. Linking glacial and future climates through an ensemble of GCM simulations. *Clim. Past* **3,** 77–87 (2007).
13. Lunt, D. J. *et al.* On the causes of mid-Pliocene warmth and polar amplification. *Earth Planet. Sci. Lett.* **321–322,** 128–138 (2012).
14. Haywood, A. M. *et al.* Are there pre-Quaternary geological analogues for a future greenhouse warming? *Phil. Trans. R. Soc. A* **369,** 933–956 (2011).
15. Edwards, T. L., Crucifix, M. & Harrison, S. P. Using the past to constrain the future: how the palaeorecord can improve estimates of global warming. *Prog. Phys. Geogr.* **31,** 481–500 (2007).
16. Crucifix, M. Does the Last Glacial Maximum constrain climate sensitivity? *Geophys. Res. Lett.* **33,** L18701 (2006).
Presents first key evidence on the state-dependence of climate sensitivity.
17. Laîné, A., Kageyama, M., Braconnot, P. & Alkama, R. Impact of greenhouse gas concentration changes on the surface energetics in the IPSL-CM4 model: regional warming patterns, land/sea warming ratio, glacial/interglacial differences. *J. Clim.* **22,** 4621–4635 (2009).
18. Otto-Bliesner, B. L. Status of CCSM4 Paleo CMIP5 Climate Simulations. http://www.cesm.ucar.edu/events/ws.2011/Presentations/Paleo/bette.pdf.
19. Held, I. M. *et al.* Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Clim.* **23,** 2418–2427 (2010).
20. Joos, F. & Spahni, R. Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years. *Proc. Natl Acad. Sci. USA* **105,** 1425–1430 (2008).
21. Köhler, P., Knorr, G., Buiron, D., Lourantou, A. & Chapellaz, J. Abrupt rise in atmospheric CO_2 at the onset of the Bølling/Allerød: in-situ ice core data versus true atmospheric signals. *Clim. Past* **7,** 473–486 (2011).
22. Hönisch, B. *et al.* The geological record of ocean acidification. *Science* **335,** 1058–1063 (2012).
23. Charney, J. G. *et al. Carbon Dioxide and Climate: A Scientific Assessment* (National Academy of Sciences, 1979).
24. Knutti, R. & Tomassini, L. Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Lett.* **35,** L09701 (2008).
25. Gregory, J. M. & Forster, P. M. Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.* **113,** D23105 (2008).
26. Soden, B. J. & Held, I. M. An assessment of climate feedbacks in coupled ocean-atmosphere models. *J. Clim.* **19,** 3354–3360 (2006).
27. Huber, M., Mahlstein, I., Wild, M., Fasullo, J. & Knutti, R. Constraints on climate sensitivity from radiation patterns in climate models. *J. Clim.* **24,** 1034–1052 (2011).
28. Huybers, P. Compensation between model feedbacks and curtailment of climate sensitivity. *J. Clim.* **23,** 3009–3018 (2010).
29. Lemoine, D. M. Climate sensitivity distributions dependence on the possibility that models share biases. *J. Clim.* **23,** 4395–4415 (2010).
30. Hansen, J. Sato, M. Kharecha, P. & von Schuckmann, K. Earth's energy imbalance and implications. *Atmos. Chem. Phys.* **11,** 13421–13449 (2011).
31. Gregory, J. M. *et al.* A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.* **31,** L03205 (2004).
32. Lambert, F. *et al.* Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452,** 616–619 (2008).
33. Winckler, G., Anderson, R. F., Fleisher, M. Q., McGee, D. & Mahowald, N. Covariant glacial-interglacial dust fluxes in the equatorial Pacific and Antarctica. *Science* **320,** 93–96 (2008).
34. Roberts, A. P., Rohling, E. J., Grant, K. M., Larrasoaña, J. C. & Liu, Q. Atmospheric dust variability from major global source regions over the last 500,000 years. *Quat. Sci. Rev.* **30,** 3537–3541 (2011).
35. Ruth, U., Wagenbach, D., Steffensen, J. P. & Bigler, M. Continuous record of microparticle concentration and size distribution in the central Greenland NGRIP ice core during the last glacial period. *J. Geophys. Res.* **108,** 4098, http://dx.doi.org/10.1029/2002JD002376 (2003).
36. Naafs, B. D. A. *et al.* Strengthening of North American dust sources during the late Pliocene (2.7 Ma). *Earth Planet. Sci. Lett.* **317–318,** 8–19 (2012).
37. Kohfeld, K. E. & Harrison, S. P. DIRTMAP: the geological record of dust. *Earth Sci. Rev.* **54,** 81–114 (2001).
38. Mahowald, N., Albani, S., Engelstaedter, S., Winckler, G. & Goman, M. Model insight into glacial-interglacial paleodust records. *Quat. Sci. Rev.* **30,** 832–854 (2011).
39. Rohling, E. J., Medina-Elizalde, M., Shepherd, J. G., Siddall, M. & Stanford, J. D. Sea surface and high-latitude temperature sensitivity to radiative forcing of climate over several glacial cycles. *J. Clim.* **25,** 1635–1656 (2012).
40. Gray, L. J. *et al.* Solar influences on climate. *Rev. Geophys.* **48,** RG4001 (2010).
41. Milankovitch, M. *Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem* (Special Publication 133, Mathematics and Natural Sciences Section, Royal Serbian Academy, Belgrade, 1941).
42. Berger, A. Support for the astronomical theory of climatic change. *Nature* **269,** 44–45 (1977).
43. Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428,** 261–285 (2004).
44. Lunt, D. J. *et al.* Earth system sensitivity inferred from Pliocene modelling and data. *Nature Geosci.* **3,** 60–64 (2010).
Presents a definition of Earth system sensitivity that includes both fast and slow processes, and its application to the Pliocene.
45. Gregory, J. & Webb, M. Tropospheric adjustment induces a cloud component in CO_2 forcing. *J. Clim.* **21,** 58–71 (2008).
46. Lüthi, D. *et al.* High-resolution CO_2 concentration record 650,000–800,000 years before present. *Nature* **453,** 379–382 (2008).
47. Loulergue, L. *et al.* Orbital and millennial-scale features of atmospheric CH_4 over the past 800,000 years. *Nature* **453,** 383–386 (2008).
48. Schilt, A. *et al.* Glacial-interglacial and millennial-scale variations in the atmospheric nitrous oxide concentration during the last 800,000 years. *Quat. Sci. Rev.* **29,** 182–192 (2010).
49. Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21,** 295–305 (2002).
50. Rohling, E. J. *et al.* Antarctic temperature and global sea level closely coupled over the past five glacial cycles. *Nature Geosci.* **2,** 500–504 (2009).
51. Bintanja, R., van de Wal, R. & Oerlemans, J. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437,** 125–128 (2005).
52. Hansen, J. *et al.* Target atmospheric CO_2: where should humanity aim? *Open Atmos. Sci. J.* **2,** 217–231 (2008).
53. Imbrie, J. & Imbrie, J. Z. Modeling the climatic response to orbital variations. *Science* **207,** 943–953 (1980).
54. Huybers, P. & Denton, G. H. Antarctic temperature at orbital timescales controlled by local summer duration. *Nature Geosci.* **1,** 787–792 (2008).
55. Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313,** 508–511 (2006).
56. Beerling, D. J. & Royer, D. L. Fossil plants as indicators of the Phanerozoic global carbon cycle. *Annu. Rev. Earth Planet. Sci.* **30,** 527–556 (2002).
57. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309,** 600–603 (2005).
58. Hönisch, B., Hemming, N. G., Archer, D., Siddall, M. & McManus, J. F. Atmospheric carbon dioxide concentration across the mid-Pleistocene transition. *Science* **324,** 1551–1554 (2009).
59. Lowenstein, T. K. & Demicco, R. V. Elevated Eocene atmospheric CO_2 and its subsequent decline. *Science* **313,** 1928 (2006).
60. Beerling, D. J. & Royer, D. L. Convergent Cenozoic CO_2 history. *Nature Geosci.* **4,** 418–420 (2011).
61. van de Wal, R. S. W., de Boer, B., Lourens, L. J., Köhler, P. & Bintanja, R. Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. *Clim. Past* **7,** 1459–1469 (2011).
Compiles CO_2 data from a variety of approaches over the past 20 million years, and condenses these into one time series.
62. Beerling, D. J., Fox, A., Stevenson, D. S. & Valdes, P. J. Enhanced chemistry-climate feedbacks in past greenhouse worlds. *Proc. Natl Acad. Sci. USA* **108,** 9770–9775 (2011).
63. de Boer, B., van de Wal, R. S. W., Lourens, L. J. & Bintanja, R. Transient nature of the Earth's climate and the implications for the interpretation of benthic $\delta^{18}O$ records. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **335–336,** 4–11 (2011).
64. Cramer, B. S., Miller, K. G., Barrett, P. J. & Wright, J. D. Late Cretaceous-Neogene trends in deep ocean temperature and continental ice volume: reconciling records of benthic foraminiferal geochemistry ($\delta^{18}O$ and Mg/Ca) with sea level history. *J. Geophys. Res.* **116,** C12023 (2011).
65. Gasson, E. *et al.* Exploring uncertainties in the relationship between temperature, ice volume and sea level over the past 50 million years. *Rev. Geophys.* **50,** RG1005 (2012).
66. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292,** 686–693 (2001).
67. Miller, K. G., Wright, J. D. & Browning, J. V. Visions of ice sheets in a greenhouse world. *Mar. Geol.* **217,** 215–231 (2005).
68. Sluijs, A. *et al.* Eustatic variations during the Paleocene-Eocene greenhouse world. *Paleoceanography* **23,** PA4216 (2008).
69. Dickens, G. R., Castillo, M. M. & Walker, J. C. G. A blast of gas in the latest Paleocene: simulating first-order effects of massive dissociation of oceanic methane hydrate. *Geology* **25,** 259–262 (1997).
70. Lourens, L. J. *et al.* Astronomical pacing of late Palaeocene to early Eocene global warming events. *Nature* **435,** 1083–1087 (2005).
71. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451,** 279–283 (2008).
72. Zeebe, R. E., Zachos, J. C. & Dickens, G. R. Carbon dioxide forcing alone insufficient to explain Palaeocene-Eocene Thermal Maximum warming. *Nature Geosci.* **2,** 576–580 (2009).
73. Huber, M. & Caballero, R. The early Eocene equable climate problem revisited. *Clim. Past* **7,** 603–633 (2011).
74. Lorius, C., Jouzel, J., Raynaud, D., Hansen, J. & Le Treut, H. The ice-core record: climate sensitivity and future greenhouse warming. *Nature* **347,** 139–145 (1990).
75. Masson-Delmotte, V. *et al.* Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints. *Clim. Dyn.* **26,** 513–529 (2006).
76. Lea, D. The 100000-yr cycle in tropical SST, greenhouse gas forcing, and climate sensitivity. *J. Clim.* **17,** 2170–2179 (2004).
77. Hansen, J. *et al.* Climate change and trace gases. *Phil. Trans. R. Soc. Lond. A* **365,** 1925–1954 (2007).
78. Bijl, P. K. *et al.* Transient Middle Eocene atmospheric CO_2 and temperature variations. *Science* **330,** 819–821 (2010).
79. Schmittner, A. *et al.* Climate sensitivity estimated from temperature reconstructions of the Last Glacial Maximum. *Science* **334,** 1385–1388 (2011).
80. Rohling, E. J. Progress in palaeosalinity: overview and presentation of a new approach. *Paleoceanography* **22,** PA3215 (2007).

81. MARGO project members. Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geosci.* **2,** 127–132 (2009).
82. Schneider von Deimling, T., Ganopolski, A., Held, H. & Rahmstorf, S. How cold was the Last Glacial Maximum? *Geophys. Res. Lett.* **33,** L14709 (2006).
83. Ballantyne, A. P., Lavine, M., Crowley, T. J., Liu, J. & Baker, P. B. Meta-analysis of tropical surface temperatures during the Last Glacial Maximum. *Geophys. Res. Lett.* **32,** L05712 (2005).
84. Hansen, J. *et al.* in *Climate Processes and Climate Sensitivity* (eds Hansen, J. & Takahashi, T.) 130–163 (Geophysical Monographs 29, American Geophysical Union, 1984).
85. Hansen, J. E. & Sato, M. in *Climate Change: Inferences from Paleoclimate and Regional Aspects* (eds Berger, A., Mesinger, F. & Šijački, D.) 21–48 (Springer, 2012).
86. Hoffert, M. I. & Covey, C. Deriving global climate sensitivity from palaeoclimate reconstructions. *Nature* **360,** 573–576 (1992).
87. Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A. C. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nature Geosci.* **3,** 27–30 (2010).
88. Covey, C., Sloan, L. C. & Hoffert, M. I. Paleoclimate data constraints on climate sensitivity: the paleocalibration method. *Clim. Change* **32,** 165–184 (1996).
89. Zachos, J. C., Stott, L. D. & Lohmann, K. C. Evolution of early Cenozoic marine temperatures. *Paleoceanography* **9,** 353–387 (1994).
90. Sloan, L. C. & Barron, E. J. A comparison of Eocene climate model results to quantified paleoclimatic interpretations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **93,** 183–202 (1992).
91. Berner, R. A. A model for atmospheric CO_2 over Phanerozoic time. *Am. J. Sci.* **291,** 339–376 (1991).
92. Freeman, K. H. & Hayes, J. M. Fractionation of carbon isotopes by phytoplankton and estimates of ancient CO_2 levels. *Glob. Biogeochem. Cycles* **6,** 185–198 (1992).
93. Cerling, T. E. Carbon dioxide in the atmosphere; evidence from Cenozoic and Mesozoic paleosols. *Am. J. Sci.* **291,** 377–400 (1991).
94. Royer, D. L., Pagani, M. & Beerling, D. J. Geobiological constraints on Earth system sensitivity to CO_2 during the Cretaceous and Cenozoic. *Geobiology* **10,** 298–310 (2012).
95. Panchuk, K., Ridgwell, A. & Kump, L. R. Sedimentary response to Paleocene-Eocene Thermal Maximum carbon release: a model-data comparison. *Geology* **36,** 315–318 (2008).
96. Borzenkova, I. I. Determination of global climate sensitivity to the gas composition of the atmosphere from paleoclimatic data. *Izv. Atmos. Ocean. Phys.* **39,** 197–202 (2003).
97. Park, J. & Royer, D. L. Geologic constraints on the glacial amplification of Phanerozoic climate sensitivity. *Am. J. Sci.* **311,** 1–26 (2011).
98. Schmidt, G. A. Climate sensitivity — how sensitive is Earth's climate to CO_2; past. *PAGES News* **20,** 11 (2012).
99. Wunsch, C. & Heimbach, P. How long to oceanic tracer and proxy equilibrium? *Quat. Sci. Rev.* **27,** 637–651 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements This Perspective arose from the first PALAEOSENS workshop in March 2011. We thank the Royal Netherlands Academy of Arts and Sciences (KNAW) for funding and hosting this workshop in Amsterdam, PAGES for their support, and J. Gregory for discussions. This study was supported by the UK-NERC consortium iGlass (NE/I009906/1), and 2012 Australian Laureate Fellowship FL120100050. D.J.B., E.J.R. and P.V. were supported by Royal Society Wolfson Research Merit Awards. A.S. thanks the European Research Council for ERC starting grant 259627, and M.H. acknowledges NSF P2C2 grant 0902882. Some of the work was supported by grant 243908 'Past4Future' of the EU's seventh framework programme; this is Past4Future contribution number 30.

Author Contributions E.J.R., A.S. and H.A.D. initiated the PALAEOSENS workshop, and led the drafting of this study together with P.K., A.S.v.d.H. and R.S.W.v.d.W. The other authors contributed specialist insights, discussions and feedback.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.J.R. (e.rohling@noc.soton.ac.uk).

PALAEOSENS Project Members E. J. Rohling[1,2], A. Sluijs[3], H. A. Dijkstra[4], P. Köhler[5], R. S. W. van de Wal[4], A. S. von der Heydt[4], D. J. Beerling[6], A. Berger[7], P. K. Bijl[3], M. Crucifix[7], R. DeConto[8], S. S. Drijfhout[9], A. Fedorov[10], G. L. Foster[1], A. Ganopolski[11], J. Hansen[12], B. Hönisch[13], H. Hooghiemstra[14], M. Huber[15], P. Huybers[16], R. Knutti[17], D. W. Lea[18], L. J. Lourens[3], D. Lunt[19], V. Masson-Demotte[20], M. Medina-Elizalde[21], B. Otto-Bliesner[22], M. Pagani[10], H. Pälike[1,23], H. Renssen[24], D. L. Royer[25], M. Siddall[26], P. Valdes[19], J. C. Zachos[27] & R. E. Zeebe[28]

Affiliations for participants: [1]School of Ocean and Earth Science, University of Southampton, National Oceanography Centre, Southampton SO14 3ZH, UK. [2]Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. [3]Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. [4]Institute for Marine and Atmospheric Research Utrecht, Utrecht University, 3584 CC Utrecht, The Netherlands. [5]Alfred Wegener Institute for Polar and Marine Research (AWI), PO Box 12 01 61, 27515 Bremerhaven, Germany. [6]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. [7]Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute–Université catholique de Louvain, Chemin du Cyclotron 2, Box L7.01.11, 1348 Louvain-la-Neuve, Belgium. [8]Department of Geosciences, 611 North Pleasant Street, 233 Morrill Science Center, University of Massachusetts, Amherst, Massachusetts 01003-9297, USA. [9]Royal Netherlands Meteorological Institute, PO Box 201, 3730 AE De Bilt, The Netherlands. [10]Department of Geology and Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520-8109, USA. [11]Potsdam Institute for Climate Impact Research (PIK), PO Box 601203, 14412 Potsdam, Germany. [12]NASA Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA. [13]Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA. [14]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. [15]Earth and Atmospheric Sciences Department, Purdue University, West Lafayette, Indiana 47907, USA. [16]Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA. [17]Institute for Atmospheric and Climate Science, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland. [18]Department of Earth Science, University of California, Santa Barbara, California 93106-9630, USA. [19]School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK. [20]LSCE (IPSL/CEA-CNRS-UVSQ), UMR 8212, LCEA Saclay, 91 191 Gif sur Yvette Cedex, France. [21]Centro de Investigación Científica de Yucatán, Unidad Ciencias del Agua, Cancún, Quintana Roo, 77500, México. [22]National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307-3000, USA. [23]MARUM, University of Bremen, Leobener Straße, 28359 Bremen, Germany. [24]Department of Earth Sciences, Faculty of Earth and Life Sciences, Free University Amsterdam, De Boelelaan 1085, NL1081HV Amsterdam, The Netherlands. [25]Department of Earth and Environmental Sciences, Wesleyan University, Middletown, Connecticut 06459, USA. [26]Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK. [27]Earth and Planetary Sciences, University of California, Santa Cruz, California 95064, USA. [28]School of Ocean and Earth Science and Technology, Department of Oceanography, University of Hawaii at Manoa, 1000 Pope Road, MSB 629 Honolulu, Hawaii 96822, USA.

doi:10.1038/nature11579

The mystery of recent stratospheric temperature trends

David W. J. Thompson[1], Dian J. Seidel[2], William J. Randel[3], Cheng-Zhi Zou[4], Amy H. Butler[5], Carl Mears[6], Albert Osso[7], Craig Long[5] & Roger Lin[5]

A new data set of middle- and upper-stratospheric temperatures based on reprocessing of satellite radiances provides a view of stratospheric climate change during the period 1979–2005 that is strikingly different from that provided by earlier data sets. The new data call into question our understanding of observed stratospheric temperature trends and our ability to test simulations of the stratospheric response to emissions of greenhouse gases and ozone-depleting substances. Here we highlight the important issues raised by the new data and suggest how the climate science community can resolve them.

The radiative effects of human emissions of ozone-depleting substances and greenhouse gases have driven marked atmospheric cooling at stratospheric altitudes[1–5]. Ozone depletion is believed to have caused the preponderance of the cooling in the lower stratosphere (around 15–25 km altitude); both ozone depletion and increases in greenhouse gases are believed to have driven the cooling in the middle and upper stratosphere (around 25–50 km altitude)[2]. Stratospheric temperature trends play an important part in allowing us to distinguish between the climate responses to natural and anthropogenic climate forcings[6]. Although less widely discussed in either scientific or policy circles, stratospheric cooling is as fundamental as surface warming as evidence of the influence of anthropogenic emissions on the climate system.

Unfortunately, observations of stratospheric temperatures are limited. The surface temperature record extends for over a century and is derived from multiple data sources[7]. In contrast, the stratospheric temperature record spans only a few decades and is derived from a handful of data sources[3,4]. Radiosonde (weather balloon) measurements are available in the lower stratosphere but do not extend to the middle and upper stratosphere[3,8]. Lidar (light detection and ranging) measurements extend to the middle and upper stratosphere but have very limited spatial and temporal sampling[3,9]. By far the most abundant observations of long-term stratospheric temperatures are derived from satellite measurements of long-wave radiation emitted by Earth's atmosphere.

The longest-running records of remotely sensed stratospheric temperatures are provided by the Microwave Sounding Unit (MSU), the Advanced Microwave Sounding Unit (AMSU), and the Stratospheric Sounding Unit (SSU). The SSU and MSU instruments were flown onboard a consecutive series of seven NOAA polar-orbiting satellites that partially overlap in time from late 1978 to 2006; the AMSU instruments have been flown onboard NOAA satellites from mid-1998 to the present day[3].

The MSU, AMSU and SSU temperature measurements do not represent temperatures at discrete height levels, but rather are representative of temperatures averaged over a continuum of altitudes described by the appropriate instrument 'weighting functions' (see, for example, Figure 2 in ref. 4). The weighting function for the highest available MSU channel (MSU channel 4) peaks in the lower stratosphere near 20 km altitude. The weighting functions for the SSU instrument peak in the middle and upper stratosphere at 25–35 km (SSU channel 1), 35–45 km (SSU channel 2), and 40–50 km (SSU channel 3).

Continuous time series of lower-stratospheric temperatures are derived by combining measurements from satellites that carried MSU instruments from 1978–2005 and AMSU instruments from 1998 to the present[3]. The lower-stratospheric MSU and AMSU data have been processed and combined by three different research groups: Remote Sensing Systems (RSS)[10], the University of Alabama-Hunstville (UAH)[11], and the NOAA Center for Satellite Applications and Research (STAR)[12]. The processing methodologies and resulting lower-stratospheric temperature data have been published extensively in the peer-reviewed literature[3,4].

Global-mean lower-stratospheric temperatures derived from the three primary stratospheric MSU products are very similar to each other (red, purple and green lines in Fig. 1d (the red, purple and green lines in Fig. 1d are reproduced in Fig. 1h to facilitate comparison with model simulations, as discussed below); the large but short-lived warmings starting in 1982 and 1991 are due to the volcanic eruptions of El Chichón and Mount Pinatubo, respectively). They are also very similar to lower-stratospheric temperatures estimated from radiosonde data[3,4]. The differences among the three MSU lower-stratospheric global-mean temperature time series are larger than those associated with separate estimates of global-mean surface temperatures[4]. And yet the differences between the MSU time series pale in comparison with those associated with the primary SSU products, as demonstrated below.

The mystery

Conflicting evidence

Continuous time series of temperatures in the middle and upper stratosphere back to 1979 are based exclusively on SSU data (the AMSU data also sample the middle and upper stratosphere but are available only since 1998). The SSU data require correction for several unique issues before they can be used for climate studies (see discussion in ref. 4). For example, (1) the SSU instrument relies critically on a cell pressure modulator of carbon dioxide to determine the emission of stratospheric radiation from different altitudes. The cells in all SSU instruments leaked with time, causing changes in the altitudes being measured; (2) the amplitude of the atmospheric thermal tides—and thus the tidal corrections between successive satellite missions—is relatively large in the middle and upper stratosphere; (3) long-term increases in atmospheric carbon dioxide influence the weighting function of the instrument; and (4) there is no overlap period between several pairs of consecutive satellites.

---

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado 80523, USA. [2]NOAA Air Resources Laboratory, College Park, Maryland 20740, USA. [3]Atmospheric Chemistry Division, NCAR, Boulder, Colorado 80307, USA. [4]NOAA/NESDIS/Center for Satellite Applications and Research , College Park, Maryland 20740, USA. [5]NOAA/NWS/NCEP/Climate Prediction Center, College Park, Maryland 20740, USA. [6]Remote Sensing Systems, Santa Rosa, California 95401, USA. [7]Department of Astronomy and Meteorology, University of Barcelona, Barcelona 08028, Spain.

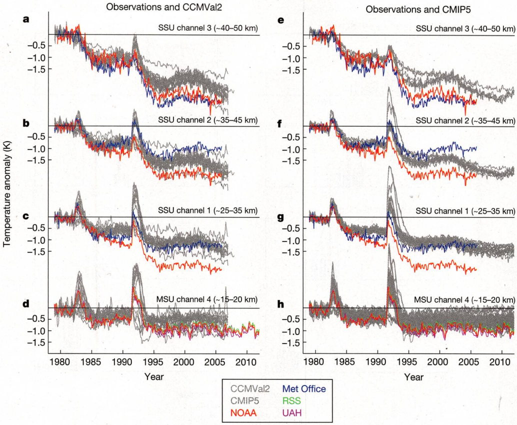


Figure 1 | Global-mean stratospheric temperature anomalies since 1979. Time series of monthly mean, global-mean stratospheric temperature anomalies are shown for the altitude ranges, data sets and model output indicated. Red, blue, green and purple lines indicate results based on observations (observations are reproduced in the left and right panels). **a–h**, Grey lines indicate results from the coupled CCM runs available through the CCMVal2 archive (**a–d**) and from the AOGCM runs available through the CMIP5 archive (**e–h**). Model runs are listed in Table 1 and were converted to SSU and MSU time series using the appropriate instrument weighting functions[9,15]. Time series are plotted so that their 1979–1982 mean anomalies are zero. Note that several CMIP5 models have poor vertical resolution at middle and upper stratospheric altitudes. For this reason, more model simulations are available at lower than at upper stratospheric levels (see Table 1).

The SSU data were originally processed for climate analysis by scientists at the UK Met Office in the 1980s[13,14]. The data were further revised in 2008 to account for variations in the satellite weighting functions over time due to changes in atmospheric composition[15]. However, the methodology used to develop the Met Office SSU product was never published in the peer-reviewed literature, and certain aspects of the original processing remain unknown. For this reason, the NOAA STAR recently reprocessed the SSU temperatures and published the full processing methodology and the resulting data in the peer-reviewed literature[16].

The new NOAA SSU data provide an invaluable independent resource for assessing the reproducibility of the original Met Office SSU data. But the new data raise more questions than they answer, because they provide a strikingly different view of recent stratospheric temperature trends (compare the red and blue lines in Fig. 1a–c; the red and blue lines in Fig. 1a–c are reproduced in Fig. 1e–g to facilitate comparison with model simulations, as discussed below). The long-term variability and trends in global-mean temperatures for the uppermost SSU channel (SSU channel 3) are relatively similar in both the Met Office and NOAA data sets. But the same cannot be said for the SSU channels that sample the middle stratosphere (SSU channels 1 and 2). The global-mean cooling in channels 1 and 2 (around 25–45 km) is nearly twice as large in the NOAA data set as it is in the Met Office data set (Figs 1 and 2)[16]. The differences between the NOAA and Met Office channels 1 and 2 global-mean time series do not arise from a discrete period of time, but rather increase from about 1985 to the end of the record[16]. The differences between the NOAA and Met Office global-mean time series shown in Fig. 1 are so large they call into question our fundamental understanding of observed temperature trends in the middle and upper stratosphere.

Disconnects between observations and models

The story is further muddled when the observations are compared with attempts to simulate the past few decades of stratospheric climate change using climate models. Two classes of climate models commonly used in simulations of past climate are coupled chemistry–climate models (CCMs) and coupled atmosphere–ocean global climate models (AOGCMs). By definition, the CCMs explicitly simulate stratospheric chemical processes, whereas the AOGCMs explicitly simulate coupled atmosphere–ocean interactions. In principle, a coupled chemistry–climate model might also simulate coupled atmosphere–ocean interactions, and vice versa. But owing to computational limitations, most current CCMs are not AOGCMs, and vice versa. A key distinction between the model classes that is pertinent to this discussion is that in general the CCMs resolve the stratosphere more fully than do the AOGCMs.

Simulations from CCMs forced with the time history of anthropogenic emissions are available via the CCM validation activity (Figs 1a–d and 2a–d, results are from the CCMVal2 project; see Table 1 and ref. 17). Between 40 and 50 km (channel 3), global-mean temperature trends from both SSU products show more cooling than is simulated by the CCMs (Figs 1a and 2a; the model temperatures are weighted by the appropriate satellite weighting functions). Between about 35 and 45 km (channel 2), the Met Office version of the SSU data suggests that the models overestimate the observed stratospheric cooling, whereas the NOAA SSU data suggest that the models underestimate it (Figs 1b and 2b). The most striking discrepancies are between about 25 and 35 km (channel 1; Figs 1c and 2c). As demonstrated in refs 4 and 18, the Met Office SSU data are in reasonable agreement with the current generation of coupled CCMs at these altitudes. But as shown in Figs 1a–d and 2a–d, the cooling

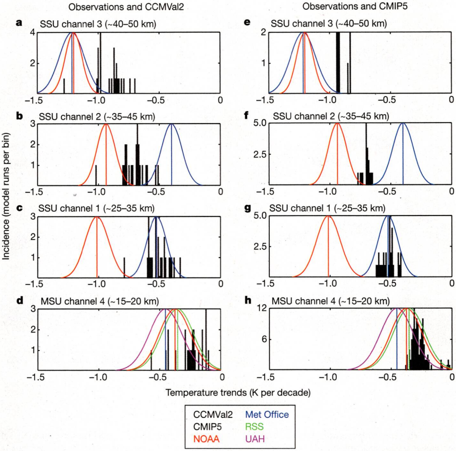


Figure 2 | Trends in global-mean stratospheric temperatures between 1979 and 2005. Trends in monthly mean, global-mean stratospheric temperatures are shown for the altitude ranges, data sets and model output indicated. Observed trends are denoted by the red, blue, green and purple vertical lines (observed trends are reproduced in the left and right panels). The normalized red, blue, green and purple probability distribution functions indicate the confidence ranges on the trend estimates, taking into account the effective number of degrees of freedom in the respective time series (for example, the 95% confidence bounds correspond to the edges of the area that spans the middle 95% of the distribution function). **a**–**h**, Black bars show the histograms of the trends from the CCM runs available through the CCMVal2 archive (**a**–**d**) and from the AOGCM runs available through the CMIP5 archive (**e**–**h**). Each temperature trend bin is 0.01 K per decade wide. The total number of model runs is given in Table 1.

in the new NOAA SSU channel 1 data is nearly twice as large as the cooling simulated by most of the CCMs.

A similar story emerges when observations of global-mean stratospheric temperature are compared with the simulations of AOGCMs prepared for the upcoming IPCC Fifth Assessment Report (Figs 1e–h and 2e–h; results are from the Coupled Model Intercomparison Project Phase 5 simulations, CMIP5; see Table 1). Most of the CMIP5 models are not coupled CCMs and have considerably less vertical resolution at stratospheric altitudes than the models archived by CCMVal2. For this reason, relatively few CMIP5 model runs include altitudes sampled by SSU channels 2 and 3.

The differences between the CMIP5 models and the observations are comparable to those noted in association with the CCMVal2 models in all SSU channels (Figs 1e–h and 2e–h). The CMIP5 models indicate considerably less cooling than both SSU products at about 40–50 km (channel 3); lie between the two SSU products at about 35–45 km (channel 2); and provide a closer fit to the Met Office SSU data than the NOAA SSU data at about 25–35 km (channel 1).

It is possible that the models are correct and that both SSU data sets are in error. But the CCMs and AOGCMs also exhibit smaller yet systematic discrepancies with observations in the lower stratosphere, which is sampled by the MSU channel 4 instrument (Figs 1d, h and 2d, h). With few exceptions, the models underestimate the amplitude of the long-term cooling in the lower stratosphere (Figs 1d, h and 2d, h) and have difficulty simulating the amplitude of the response to the eruptions of El Chichón and Mount Pinatubo there (Fig. 1d, h). Previous studies have reported close agreement between trends in the MSU channel 4 data and in CCMVal2 simulations[18], but those trend comparisons were done between observations of MSU channel 4 temperature and model output at specific height levels (that is, the model trends were shown as a function of height and not averaged over the MSU channel 4 weighting function; see figure 2 in ref. 18, for example).

The latitudinal profiles of the trends from the different SSU data sources are also remarkably different from those simulated by the current generation of CCMs (Fig. 3). The Met Office SSU data suggest that the cooling of the past few decades was relatively uniform with latitude (blue lines in Fig. 3a–c). In contrast, the NOAA SSU data suggest that the largest stratospheric cooling occurred at tropical latitudes, particularly between 25 km and 45 km (red lines in Fig. 3a–c). The differences between the Met Office and NOAA global-mean stratospheric temperature trends clearly derive primarily from tropical latitudes. The tropical stratospheric cooling indicated by the models is noticeably weaker than that indicated by the NOAA SSU data in the middle and upper stratosphere (Fig. 3a–c), and is generally weaker than that indicated by all MSU channel 4 products in the lower tropical stratosphere (Fig. 3d).

What might cause cooling in the tropical stratosphere? The radiative effects of increasing carbon dioxide are modest below 40 km altitude[2]. Rather, at altitudes sampled by SSU channel 1, long-term tropical cooling is most likely to result from either anomalous rising motion, which decreases air temperature through expansion, or *in situ* ozone depletion, which decreases temperature by reducing the absorption of short-wave radiation. The two processes are closely related: rising motion leads to

Table 1 | Model runs used in this study

CMIP5 model runs	CCMVal2 model runs
CanESM2* (5)	AMTRAC3
CCSM4 (6)	CCSRNIES
CSIRO-Mk3.6.0 (10)	CMAM (3)
FGOALS-s2 (3)	EMAC
GFDL-CM3* (5)	LMDZrepro (3)
GFDL-ESM2G (1)	MRI (4)
GFDL-ESM2M (1)	NIWA SOCOL
GISS-E2-H (15)	SOCOL (3)
GISS-E2-R (16)	ULAQ
HadCM3 (10)	UMSLIMCAT
HadGEM2-CC** (3)	WACCM (4)
HadGEM2-ES (4)	
INMCM4 (1)	
IPSL-CM5A-LR (5)	
IPSL-CM5A-MR (1)	
IPSL-CM5B-LR (1)	
MIROC4h* (3)	
MIROC5 (4)	
MIROC-ESM*** (3)	
MIROC-ESM-CHEM*** (1)	
MPI-ESM-LR*** (3)	
MPI-ESM-P*** (2)	
MRI-CGCM3** (5)	
NorESM1-M (3)	
NorESM1-ME (1)	

Numbers in parentheses indicate the number of ensemble members. The number of asterisks indicates at which level the model temperature data was used based on which levels are available in model output. No asterisks means used only in MSU channel 4 (any model with no output at 1 hPa). *Used in MSU channel 4 and SSU channel 1 (any model with output at 1 hPa). **Used in MSU channel 4, SSU channels 1 and 2 (any model with output at pressures below 1 hPa). ***Used in all channels (any model with output at pressures below 0.1 hPa). Model nomenclature is provided in ref. 17 and the IPCC Fifth Assessment Report (http://www.ipcc.ch/).

decreases in ozone in the lower stratosphere through the vertical transport of low-ozone air from lower altitudes. The two processes are also both potentially implicated in recent stratospheric climate change.

Rising motion in the tropical stratosphere occurs as part of the large-scale, equator-to-pole stratospheric mass circulation. Most coupled CCMs suggest that increasing greenhouse gases accelerate the stratospheric mass circulation[19–24]. Such an acceleration is expected to be marked by decreases in tropical stratospheric ozone and temperatures (particularly in the lower stratosphere), and observations suggest that both changes are occurring. Ground-based and satellite measurements suggest that tropical lower-stratospheric ozone has decreased over the past few decades at a rate comparable to that predicted by the CCMs[25,26], radiosonde data suggest that tropical lower-stratospheric temperatures have decreased since 1979[3,27–29] and the NOAA SSU data suggest that such tropical cooling extends to the middle and upper stratosphere (Fig. 3). (In principle, the acceleration of the mass circulation should also be marked by increases in temperatures and ozone concentrations in the extratropical stratosphere owing to the anomalous downward motion there, but the effects of the mass circulation at extratropical latitudes are opposed by the effects of polar stratospheric chemical ozone depletion in the Southern Hemisphere—the 'Antarctic ozone hole'—and masked by naturally high levels of year-to-year climate variability in the Northern Hemisphere.)

If the new NOAA SSU data are correct, they suggest that the stratospheric mass circulation is accelerating at a rate considerably higher than that predicted by the CCMs, at least in the middle and upper stratosphere (that is, at the altitudes sampled by the SSU instrument). Again, it is possible that the models are correct and that the SSU data are in error. But the fact that the discrepancies between the magnitudes of the simulated and observed cooling in the tropical stratosphere extend to MSU channel 4, which samples the lower stratosphere and exhibits trends that are fairly reproducible from one data set to the next (Figs 1d, h, 2d, h and 3d), suggest that model uncertainties should not be discounted.

Moving forward to resolve the mystery

Are the models missing a key aspect of stratospheric climate change? Or is there an error in the newly processed NOAA data? Which SSU data set is correct? Or are both in error?

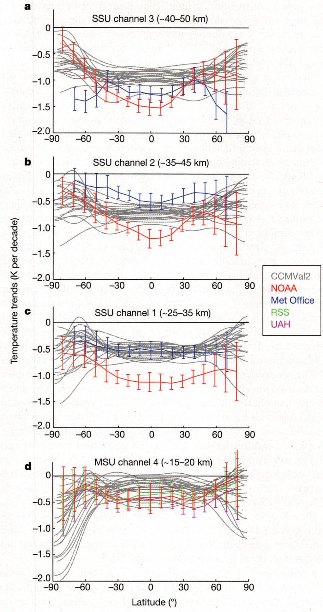


Figure 3 | The north–south structure of zonal-mean stratospheric temperature trends between 1979 and 2005. a–d, Trends in monthly mean, zonal-mean stratospheric temperatures are shown for the altitude ranges, data sets and model output indicated. Error bars approximate the 95% confidence bounds.

The latitudinal structure of stratospheric temperature trends is influenced by both radiative processes and variability in the stratospheric mass circulation. As noted above, the predicted acceleration of the stratospheric mass circulation[19–24] should lead to enhanced stratospheric cooling at tropical latitudes, and such cooling is found in the middle and upper stratosphere in both the NOAA SSU data and the CCMVal2 simulations (Fig. 3). However, the magnitudes of the predicted acceleration are not well constrained by theory, and the latitudinal structure of the trends exhibits considerable variability from data set to data set and—to a lesser extent—model to model[30] (Fig. 3). Trends in the stratospheric mass circulation are difficult both to detect and to predict.

In contrast, trends in global-mean stratospheric temperature are relatively simple to constrain quantitatively. The influence of the stratospheric mass circulation on temperature trends is negligible in the global-mean temperature because the regions of upward and downward motion average out. Thus trends in global-mean stratospheric temperatures are driven almost entirely by the radiative effects of changes in stratospheric composition, primarily increases in carbon dioxide and changes in ozone concentrations, but also changes in the concentrations of water vapour, aerosols and other trace gases. If the NOAA SSU data are correct, then both the CCMVal2 and CMIP5 models are presumably missing key changes in stratospheric composition.

What might give rise to the discrepancies between observed and simulated global-mean stratospheric temperatures highlighted here? The long-term increases in stratospheric concentrations of carbon dioxide are probably well constrained by observations and in models owing to the fact that carbon dioxide is largely inert and thus well mixed in the atmosphere. Simulations of stratospheric water vapour trends and their effects on temperature vary considerably from model to model[18]. But the effects on temperature of stratospheric water vapour trends are much more important in the lower stratosphere than they are in the middle and upper stratosphere[2]. Therefore, uncertainties in simulated stratospheric water vapour trends may contribute to the discrepancies between simulated and observed temperature trends in the lower stratosphere, but they seem unlikely to contribute significantly to the discrepancies in the middle and upper stratosphere. Trace gases such as nitrous oxide, methane and fluorinated greenhouse gases are not believed to have had a pronounced effect on trends in the middle and upper stratosphere[2,18]. And multiple observational sources suggest that the overall trends in stratospheric aerosols were very small over the SSU period (about 1979–2005)[31]. Hence, the pronounced discrepancies between simulated and observed global-mean stratospheric temperature trends are most probably due to one of the following two possibilities.

(1) The observations may be in error. The MSU channel 4 temperature record is robust from one data set to the next, so we consider it to be unlikely that uncertainties in the MSU channel 4 data can account for the discrepancies between modelled and observed lower stratospheric temperatures shown here. Uncertainties in middle and upper stratospheric temperatures derived from the SSU instrument are much larger.

(2) The simulated ozone trends may be in error. The observed and simulated global-mean ozone trends are very similar in both the middle and upper stratosphere[26]. We therefore consider it to be unlikely that the differences between modelled and observed temperature trends in the middle and upper stratosphere can be explained by differences in ozone trends at these altitudes. Uncertainties in ozone depletion in the lower stratosphere[32] may help to account for the discrepancies between modelled and observed trends in temperatures there.

How might the climate community resolve the mysteries raised by the new SSU data? First, the methodology used to generate the original Met Office SSU data remains undocumented and so the climate community are unable to explain the large discrepancies between the original Met Office and NOAA SSU products highlighted here. The World Climate Research Programme's Stratospheric Temperature Trends Assessment Panel (of which several authors of this study are members) has encouraged the scientists who generated the original Met Office data set to publish the methodology, but they are now retired. We encourage the Met Office to allocate resources towards the recovery and publication of as much of the original SSU metadata as possible.

Second, the SSU data should be processed by at least one additional independent research group. Similar controversies regarding surface and tropospheric temperature changes over the past few decades have motivated tests of the reproducibility of trend estimates. Other key data sources are now routinely vetted, processed and published by a number of research organizations: scientists have produced at least three independent MSU temperature products, five independent radiosonde temperature products and five global surface temperature products for climate research (see discussion in refs 4 and 7). The SSU data have been processed by only two independent research groups, and published by only one.

Third, the amplitudes of the observed stratospheric ozone depletion should be critically assessed in all available data sources, discrepancies between simulated and observed variability in stratospheric ozone should continue to be explored, and remotely sensed observations used to estimate stratospheric ozone depletion should be processed by independent research groups (for example, as done for MSU channel 4 temperatures). The World Meteorological Organization and International Ozone Commission are supporting an effort to critically evaluate ozone profile trends based on remotely sensed and *in situ* measurements. It remains to be seen whether revised estimates of stratospheric ozone depletion are large enough to account for the discrepancies between observed and modelled stratosphere temperature trends highlighted here.

Finally, to avoid a continuation of the current perplexing and frustrating situation, it is imperative that stratospheric altitudes are included in future climate reference data networks. The Global Climate Observing System (GCOS)—a project overseen by the World Meteorological Organization, the United Nations Environment Program, and other international bodies—is currently developing a 'reference' upper-air network consisting of around 30–40 ground-based stations that will be used to constrain the numerous atmospheric observations used in climate research (the GCOS Reference Upper-Air Network; GRUAN[33]). Other than these incipient GRUAN observations, there are currently no reference temperature data at stratospheric altitudes. The GRUAN effort is essential for assessing future stratospheric climate change without the ambiguities we currently face.

Received 8 May; accepted 12 September 2012.

1. Ramaswamy, V. *et al.* Stratospheric temperature trends: observations and model simulations. *Rev. Geophys.* **39,** 71–122 (2001).
2. Shine, K. P. *et al.* A comparison of model-simulated trends in stratospheric temperatures. *Q. J. R. Meteorol. Soc.* **129,** 1565–1588 (2003).
3. Randel, W. J. *et al.* An update of observed stratospheric temperature trends. *J. Geophys. Res.* **114,** D02107 (2009).
4. Seidel, D. J., Gillett, N. P., Lanzante, J. R., Shine, K. P. & Thorne, P. W. Stratospheric temperature trends: our evolving understanding. *Wiley Interdisc. Rev. Clim. Change* **2,** 592–616 (2011).
5. Forster, P. M. *et al.* in *Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project* Report No. 52, Ch. 4 (World Meteorological Organization, 2011).
6. Hansen, J. *et al.* Forcings and chaos in interannual to decadal climate change. *J. Geophys. Res.* **102,** 25679–25720 (1997).
7. Trenberth, K. E. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) Ch. 3 (Cambridge Univ. Press, 2007).
8. Lanzante, J., Klein, S. & Seidel, D. J. Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities and MSU comparisons. *J. Clim.* **16,** 241–262 (2003).
9. Keckhut, P. *et al.* Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change. *J. Environ. Monit.* **6,** 721–733 (2004).
10. Mears, C. A. & Wentz, F. J. Construction of the Remote Sensing Systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders. *J. Atmos. Ocean. Technol.* **26,** 1040–1056 (2009).
11. Christy, J. R., Spencer, R. W., Norris, W. B., Braswell, W. D. & Parke, D. E. Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures. *J. Atmos. Ocean. Technol.* **20,** 613–629 (2003).
12. Zou, C.-Z. *et al.* Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses. *J. Geophys. Res.* **111,** D19114 (2006).
13. Nash, J. & Forrester, G. F. Long-term monitoring of stratospheric temperature trends using radiance measurements obtained by the TIROS-N series of NOAA spacecraft. *Adv. Space Res.* **6,** 37–44 (1986).
These authors pioneered the use of infrared radiances from the SSU instrument for climate studies and produced the first SSU data set.
14. Nash, J. Extension of explicit radiance observations by the Stratospheric Sounding Unit into the lower stratosphere and lower mesosphere. *Q. J. R. Meteorol. Soc.* **114,** 1153–1171 (1988).
15. Shine, K. P., Barnett, J. J. & Randel W. J. Temperature trends derived from Stratospheric Sounding Unit radiances: the effect of increasing CO_2 on the weighting function. *Geophys. Res. Lett.* **35,** L02710 (2008).
These authors were the first to quantify the effect of increasing carbon dioxide on the SSU weighting function, and their work highlighted the need to revisit and reprocess the SSU data set of ref. 13.
16. Wang, L., Zou, C.-Z. & Qian, H. Construction of stratospheric temperature data records from stratospheric sounding units. *J. Clim.* **25,** 2931–2946 (2012).
These authors provided the first full reprocessing of the original SSU radiances, and their findings have raised serious questions regarding our understanding of stratospheric temperature trends.
17. *SPARC Report on the Evaluation of Chemistry-Climate Models* (eds Eyring, V., Shepherd, T. G. & Waugh, D. W.) SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526 http://www.sparc-climate.org (SPARC, 2010).
18. Forster, P. M. *et al.* Evaluation of radiation scheme performance within chemistry climate models. *J. Geophys. Res.* **116,** D10302 (2011).
19. Butchart, N. *et al.* Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation. *Clim. Dyn.* **27,** 727–741 (2006).

20. Garcia, R. R. & Randel, W. J. Acceleration of the Brewer-Dobson circulation due to increases in greenhouse gases. *J. Atmos. Sci.* **65,** 2731–2739 (2008).
21. Butchart, N. *et al.* Chemistry-climate model simulations of 21st century stratospheric climate and circulation changes. *J. Clim.* **23** (2010).
22. McLandress, C. & Shepherd, T. G. Simulated anthropogenic changes in the Brewer-Dobson circulation, including its extension to high latitudes. *J. Clim.* **22,** 1516–1540 (2009).
23. Shepherd, T. G. & McLandress, C. A robust mechanism for strengthening of the Brewer-Dobson circulation in response to climate change: critical-layer control of subtropical wave breaking. *J. Atmos. Sci.* **68,** 784–797 (2011).
24. Garny, H., Dameris, M., Randel, W. J., Bodeker, G. E. & Deckert, R. Dynamically forced increase of tropical upwelling in the lower stratosphere. *J. Atmos. Sci.* **68,** 1214–1233 (2011).
25. Randel, W. J. & Thompson, A. M. Interannual variability and trends in tropical ozone derived from SAGE II satellite data and SHADOZ ozonesondes. *J. Geophys. Res.* **116,** D07303 (2011).
26. Douglas, A. *et al.* in *Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project* Report No. 52, Ch. 2 (World Meteorological Organization, 2011).
27. Free, M. *et al.* Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): a new data set of large-area anomaly time series. *J. Geophys. Res.* **110,** D22101 (2005).
28. Thompson, D. W. J. & Solomon, S. Recent stratospheric climate trends as evidenced in radiosonde data: global structure and tropospheric linkages. *J. Clim.* **18,** 4785–4795 (2005).
29. Young, P. J. *et al.* Changes in stratospheric temperatures and their implications for changes in the Brewer Dobson Circulation, 1979–2005. *J. Clim.* **25,** 1759–1772 (2012).
30. Wang, L. & Waugh, D. W. Chemistry-climate model simulations of recent trends in lower stratospheric temperature and stratospheric residual circulation. *J. Geophys. Res.* **117,** D09109 (2012).
31. Deshler, T. A review of global stratospheric aerosol: measurements, importance, life cycle, and local stratospheric aerosol. *Atmos. Res.* **90,** 223–232 (2008).
32. Solomon, S., Young, P. J. & Hassler, B. Uncertainties in the evolution of statospheric ozone and implications for recent temperature changes in the tropical lower stratosphere. *Geophys. Res. Lett.* **39,** L17706 (2012).
33. Seidel, D. J. *et al.* Reference upper-air observations for climate: rationale, progress, and plans. *Bull. Amer. Meteor. Soc.* **90,** 361–369 (2009).

Acknowledgements We thank K. Shine, M. Dameris, M. Free and R. Saunders for suggestions and comments on the manuscript. D.W.J.T. is supported by the National Science Foundation Climate Dynamics Program under budget number AGS-0936255. We acknowledge the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling and the WCRP CCM Validation project (CCMVal), which are responsible for archiving the CMIP5 and CCMVal2 output, respectively. We also thank the climate modelling groups for producing and making available their model output as listed in Table 1. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Author Contributions D.W.J.T. led the analyses and the writing of the text. D.J.S., W.J.R. and C.-Z.Z. contributed to the text and analysis design and provided guidance on all aspects of the study. A.H.B. provided the CMIP5 output, assisted with the analyses of the model output, and provided advice on the text. C.M. provided expertise on the data sets used in the study and advice on the text. A.O. assisted with the analyses and data processing and provided advice on the text. C.L. and R.L. provided advice on the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.J.T. (davet@atmos.colostate.edu).

doi:10.1038/nature11678

The global pattern of trace-element distributions in ocean floor basalts

Hugh St C. O'Neill[1] & Frances E. Jenner[1,2]

The magmatic layers of the oceanic crust are created at constructive plate margins by partial melting of the mantle as it wells up. The chemistry of ocean floor basalts, the most accessible product of this magmatism, is studied for the insights it yields into the compositional heterogeneity of the mantle and its thermal structure. However, before eruption, parental magma compositions are modified at crustal pressures by a process that has usually been assumed to be fractional crystallization. Here we show that the global distributions of trace elements in ocean floor basalts describe a systematic pattern that cannot be explained by simple fractional crystallization alone, but is due to cycling of magma through the global ensemble of magma chambers. Variability in both major and incompatible trace-element contents about the average global pattern is due to fluctuations in the magma fluxes into and out of the chambers, and their depth, as well as to differences in the composition of the parental magmas.

Magmas parental to ocean floor basalts (OFBs) are traditionally considered to evolve by fractional crystallization along 'liquid lines of descent', producing the oceanic crust with ~4–5 km of gabbro underlying 1–1.5 km of basalt and dolerite[1]. The major-element compositions of OFB scatter around the low-pressure olivine+plagioclase+clinopyroxene (ol+pl+cpx) cotectic (the hypersurface in compositional space where multiple solid phases will crystallize at the same time from a single liquid), consistent with their evolution by fractional crystallization of these phases (Fig. 1). The scatter is attributed to (1) compositional variation in the source, (2) differences in extent of melting reflecting variations in mantle potential temperature[2–4], and (3) the proportion of the phases on the cotectic, which depends on both pressure of crystallization and magma composition[5–7].

However, the abundances of highly incompatible trace elements increase with indicators of evolution like MgO content ([MgO]) more rapidly than can be explained by fractional crystallization[8,9]. Although these early studies considered only local trends, the same phenomenon appears when the chemistry of OFB is viewed from a global perspective. Consider three rare earth elements (REEs), representative of incompatible trace elements in general, whose contents from two global databases of OFB[10,11] are plotted against [MgO] in Fig. 1d–f. Above ~5.5 wt% MgO, the incompatible trace elements are distributed approximately log-normally about a line of log[M] versus [MgO], where [M] is the content of M in the melt. The distribution of each M is characterized by three entities: first, the intercept, representing an average parental content, $[\mathrm{M}]_o$, taken here at [MgO] = 10 wt%; second, the slope by which the log-mean content changes with [MgO], that is, d(log[M])/d[MgO]; and third, variability about the slope. To test whether these three entities remain approximately constant with [MgO], we plot the intercepts, slopes and variabilities for Na, P, K and Ti in bins of 1 wt% [MgO] for 9,050 samples previously analysed by electron microprobe[12] (Fig. 2a–c). The large number of samples is mandatory for obtaining precise values over such short segments. All three entities remain nearly constant from 5.5 to 8.5 wt% MgO, which covers 85% of the glasses.

Plots of intercepts, slopes and variabilities of all the REEs versus ionic radii reveal systematic patterns for all three entities (Fig. 2d–f). Crystal/melt partition coefficients for REEs correlate well with ionic radii[13], with approximately a parabolic relationship, indicating that crystal/melt partitioning controls both the intercepts and slopes. Also apparent is the systematic nature of the variability, confirming previous observations[14,15].

The linearity of log([M]) with [MgO] might appear consistent with fractional crystallization, because [M] varies with the degree of crystallization F and the bulk solids/melt partition coefficient D_{M} according to the Rayleigh equation:

$$\log\,[\mathrm{M}]/[\mathrm{M}]_o = (D_{\mathrm{M}} - 1)\log F \tag{1}$$

and, as shown in Fig. 1c, the relationship between F and MgO is well approximated along the ol+pl±cpx cotectics by:

$$\log F = -1.245 + 0.132[\mathrm{MgO}] \tag{2}$$

However, fractional crystallization predicts that the slopes d(log[M])/d[MgO]) should approach a theoretical limit of −0.132 asymptotically as D_{M} goes to 0. For the REEs, for example, the slopes expected for simple fractional crystallization are shown in Fig. 2e. The actual slopes of the light REEs exceed the theoretical limit (La by nearly a factor of two), and the plot of slope versus REE ionic radius (Fig. 2e) shows no sign of the expected asymptotic behaviour. The same phenomenon is found for a suite of highly enriched basalts from Macquarie Island[16].

Trends for the other 12 trace refractory lithophile elements (RLEs) can be compared to the REEs by plotting intercepts versus slopes (Fig. 3a). The other trace RLEs with incompatibilities within the range covered by the REEs fall on the same parabola, with the exception of Sr, which plots to the right of the REEs. Sc, the least incompatible of the trace RLEs, falls off the extension of the REE array, probably because a significant proportion is retained in the mantle residue during melting. The most incompatible RLEs (Ba, Nb, Ta and Th) define a

---

[1] Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 0200, Australia. [2] Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road Northwest, Washington DC 20015, USA.

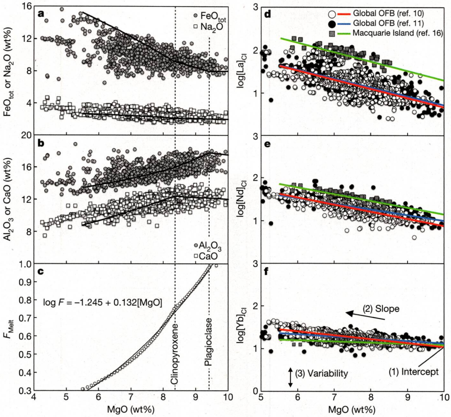


Figure 1 | Global trends in ocean floor basalt (OFB) glasses. **a–c**, Major-element trends in the crystallization of OFBs, illustrated using the global data sets of refs 10 and 11. Major-element oxides are plotted versus MgO, the oxide that is most sensitive to the degree of fractional crystallization; **a**, FeO$_{tot}$ and Na$_2$O; **b**, Al$_2$O$_3$ and CaO. The average global trends (solid curves) were modelled using PETROLOG software[42] and models for ol, pl and cpx[43]. The representative parental composition at 10 wt% MgO is taken (in wt%) as: SiO$_2$ 49, TiO$_2$ 1, Al$_2$O$_3$ 17.5, FeO$_{tot}$ 8, MgO 10, CaO 12, Na$_2$O 2, MnO, K$_2$O and H$_2$O 0.1, and P$_2$O$_5$ 0.05. Pressure is 0.3 GPa and relative oxygen fugacity is set by the quartz–fayalite–magnetite oxygen buffer. Some of the mismatch for FeO$_{tot}$ may reflect problems with traditional estimates of MORB Fe^{3+}/Fe^{2+} (ref. 44). The average mass fractions of phases crystallizing along the ol+cpx+pl cotectic are: $f^{ol} = 0.06, f^{cpx} = 0.44, f^{pl} = 0.49$. **c**, The fraction crystallized (F) is shown as a function of MgO and parameterized using an exponential relationship (solid curve). **d–f**, Variation in three REEs (**d**, La; **e**, Nd; **f**, Yb) with MgO using two global databases of OFBs[10,11] and highly enriched basalts from a single locality, Macquarie Island[16]. Elements are scaled by normalizing to CI chondrites. The lines (between 5.5 and 10 wt% MgO) are least-squares fits to the equation: $\log[\mathrm{M}]/[\mathrm{M}]_{\mathrm{CI}} = \log[\mathrm{M}]_{\mathrm{o}}/[\mathrm{M}]_{\mathrm{CI}} + \mathrm{d}(\log[\mathrm{M}])/\mathrm{d}[\mathrm{MgO}] \times ([\mathrm{MgO}] - 10)$, assuming equal weighting of log[M] and treating [MgO] as an independent variable. Each trace-element distribution is characterized by three attributes: (1) the mean of the log-normal distribution taken at 10 wt% MgO, called 'intercept'; (2) the slope of the mean of the log-normal distribution with MgO, d(log[M])/d[MgO], taken as constant; and (3) the variability, defined as: $\sigma(\log[\mathrm{M}]) = \left(\frac{\sum(\log[M] - \log[M]_{\mathrm{calc}})^2}{n}\right)^{1/2}$, where $[\mathrm{M}]_{\mathrm{calc}}$ is the average concentration at that MgO content, calculated from the intercept and slope.

minimum limiting slope, which we propose corresponds to $D_{\mathrm{M}} \approx 0$. Intercepts for the cosmochemically siderophile and volatile elements (SVEs) cannot be compared with the RLEs without introducing circular arguments, because their normalizations would assume a relationship to the RLEs[17]. The exception is Pb, whose normalization to the RLEs is constrained by U–Th–Pb isotopic systematics[17]. Because neither the slopes nor the variabilities depend on normalization, these entities for the SVEs are included in Fig. 3b. Several SVEs share the minimum limiting slope: Rb, Cs and W, and also Cl and Br from the Macquarie Island suite[18]. There is a striking negative correlation ($R = -0.95$) between intercepts and variabilities.

The model

Although basalts comprise only a minor part of the magmatic output that forms the oceanic crust (most of the parental magma crystallizes as gabbro), it has been commonplace to assume that they preserve relative abundances of the incompatible trace elements, because simple fractional crystallization should not change incompatible-element ratios at plausible degrees of crystallization (see equation (1); see also Fig. 2e). But it has been emphasized[19,20] that repeated replenishment of crustal magma chambers followed by crystallization negates this assumption. The replenished magma inherits part of its composition from previous cycles, causing incompatible elements to build up relative to compatible ones, which are preferentially removed by crystallization. If the replenishing magma has constant composition, and the fractions of the magma chamber tapped, ϕ_{T}, and crystallized, ϕ_{X}, in each cycle are constant, magma chambers of constant mass reach a steady state after a sufficient number of cycles, in which, for each element M, the flux in is balanced by the flux out:

$$[\mathrm{M}]_{\mathrm{o}}(\phi_{\mathrm{T}} + \phi_{\mathrm{X}}) = [\mathrm{M}]_{\mathrm{T}}\phi_{\mathrm{T}} + [\mathrm{M}]_{\mathrm{X}}\phi_{\mathrm{X}} \qquad (3)$$

where $[\mathrm{M}]_{\mathrm{o}}$ is the content of M in the replenishing (or parental) magma, $[\mathrm{M}]_{\mathrm{T}}$ is the content of M in the magma tapped to make basalts and dolerites, and $[\mathrm{M}]_{\mathrm{X}}$ is the content of M in the crystals. The idea of magma chambers undergoing cycles of replenishment and tapping is an attractive proposition for mid-ocean ridges. These structures sit athwart rifting zones and continually receive magma from decompression melting; they distribute the magma, with fraction ϕ_{T} to basalt and dolerite, and ϕ_{X} to cumulate gabbros, to form the crust on top of the diverging plates.

Individual magma chambers are unlikely to reach steady state because of changes in the composition of replenishing magmas, or

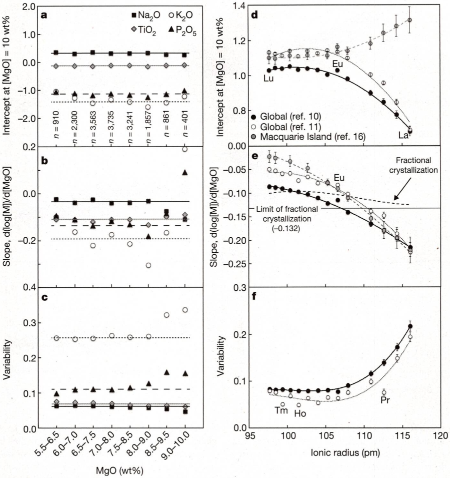


Figure 2 | Homogeneous relations among minor elements (Na, Ti, K and P) and rare earth elements (REEs) in OFB glasses. a–c, Intercepts (**a**), slopes (**b**) and variabilities (**c**) of Na, P, K and Ti from the Smithsonian abyssal volcanic glass data file[12]. The number of samples (n) in each overlapping 1 wt% MgO bin is given in **a**. P and K are near the limits of detection (especially at low contents in high-MgO samples), hence analytical uncertainty and pixelation of data[45] compromise the reliability of these data. Each characteristic is nearly constant in each MgO bin, with the exception of the two highest MgO bins, where steady state is probably not reached. The horizontal lines are the global averages. **d–f,** Intercepts (**d**), slopes (**e**) and variabilities (**f**) of REEs against ionic radius, using two global databases of OFBs[10,11] and samples from Macquarie Island[16]. The intercepts and slopes fall on parabolas, indicating control by crystal/melt partition coefficients. The slopes calculated for simple fractional crystallization using the phase proportions on the cotectic for the model in Fig. 1, and crystal/melt partition coefficients from Supplementary Table 1, are shown in **e**. The slopes for the light REEs are greater than simple fractional crystallization can produce. Eu occurs partly in 2+ as well as 3+ oxidation state, making it more compatible in plagioclase during low-pressure evolution (higher slope) compared to the other REEs. The higher intercept reflects a positive Eu anomaly in the mantle source[10]. The extra scatter in the data set of ref. 11 is due to the monoisotopic REEs (Pr, Tb, Ho and Tm) not being reported in some subsets. The curved lines are empirical fits to quadratic functions of the ionic radii.

in their size or in ϕ_T or ϕ_X, but one can envisage averages over many cycles and many chambers. Such averaging allows the global ensemble of magma chambers to be described by the average $\overline{\phi}_X$ and $\overline{\phi}_T$, replenished by magma with average contents $\overline{[M]}_o$, and fractionated with average bulk crystal/melt partition coefficients $\overline{D}_M$. The global average content of M tapped as basalt or dolerite is $\overline{[M]}_T$. The phases crystallizing to form the cumulate are ol, pl and cpx. Bulk partition coefficients, D_M, are given by:

$$D_M = D_M^{pl} f^{pl} + D_M^{cpx} f^{cpx} + D_M^{ol} f^{ol} \tag{4}$$

where D_M^{pl} is the plagioclase/melt partition coefficient (similarly for D_M^{cpx} and D_M^{ol}), and f^{pl} is the mass fraction of plagioclase in the crystallizing mineral assemblage (similarly for f^{cpx} and f^{ol}). The mass fractions are related by $f^{pl} + f^{cpx} + f^{ol} = 1$, and similarly for their global averages ($\overline{f}^{ol}$ and so on). Accessory phases with $f < 0.01$ need to be included for a few elements: spinel for Cr, and sulphide for the chalcophile elements (S, Se, Cu, Ag and probably Ni and Pb).

Elements known to be highly incompatible (M_H), but which otherwise have different chemical properties (Rb, Cs, Ba, Th, Nb, Ta, W, Cl and Br) share the same limiting minimum slopes d(log[M_H])/d[MgO] of −0.26 (Fig. 3). This observation provides the key to deciphering the global patterns, because we can assign to these elements $D_{M_H} = 0$, hence $[M_H]_X$ is zero, and, regardless of whether crystallization is fractional or equilibrium, equation (3) reduces to:

$$\frac{[M_H]_T}{[M_H]_o} = 1 + \frac{\overline{\phi}_X}{\overline{\phi}_T} \tag{5}$$

The linear relationship between log[M_H] and [MgO] can be written:

$$\begin{aligned}\log\left(\frac{[M_H]_T}{[M_H]_o}\right) &= \mathrm{d}(\log[M_H])/\mathrm{d}[MgO] \times (\overline{[MgO]}_T - \overline{[MgO]}_o)\\ &= \log\left(1 + \frac{\overline{\phi}_X}{\overline{\phi}_T}\right)\end{aligned} \tag{6}$$

It is then necessary to describe how $[MgO]_T$ varies with ϕ_X and ϕ_T. It has been noted[21] that if the replenishing magma is mixed with inherited magma, the chamber is then tapped, and subsequently the

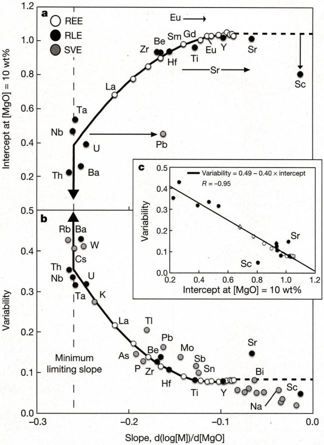


Figure 3 | Systematic relations between intercepts at 10% MgO, slopes and variabilities in OFB glasses. **a**, Intercepts versus slopes for REEs, other RLEs and Pb. **b**, Variabilities versus slopes for REEs, other RLEs and some SVEs. Note the minimum limiting slope of −0.26 reached by the most incompatible elements (Nb, Ta, Ba, Rb, Cs, W). Sr and Eu plot shifted to the right relative to REEs of similar incompatibility during mantle melting (Sr is similar to Nd, Eu falls between Sm and Gd) because of their compatibility in plagioclase; the same may explain the Pb anomaly. Sc has a lower intercept at 10 wt% MgO because of its retention in residual mantle after melt extraction due to its compatibility in orthopyroxene and olivine. Solid curves are least-squares fits to REEs only, with their extrapolations to more compatible behaviour shown as the horizontal dashed lines. **c**, Variabilities versus intercepts, with best-fit line showing the strong negative correlation ($R = -0.95$) between these entities.

remaining magma evolves by fractional crystallization before the next cycle of replenishment, $[M]_X$ is related to $[M]_T$ through the usual equation for fractional crystallization (equation (1)), which, on substitution into equation (3), gives $[M]_T$ at steady state:

$$\frac{[M]_T}{[M]_o} = \frac{(\phi_X + \phi_T)}{1 + \phi_T - (1 - \phi_X)^{D_M}} \qquad (7)$$

This equation differs from that in refs 19 and 20, which assumed fractional crystallization would follow replenishment before tapping. Applying equation (7) to MgO gives:

$$\overline{[\mathrm{MgO}]}_T = \frac{\overline{[\mathrm{MgO}]}_o(\bar{\phi}_X + \bar{\phi}_T)}{1 + \bar{\phi}_T - (1 - \bar{\phi}_X)^{\bar{D}_{\mathrm{Mg}}}} \qquad (8)$$

Equations (6) and (8) have two unknowns, $\bar{\phi}_X$ and $\bar{\phi}_T$, that are functions of the global average MgO content of the tapped magma at steady state, $\overline{[\mathrm{MgO}]}_T$. Real solutions have $0 < \bar{\phi}_X, \bar{\phi}_T < 1$ and $(\bar{\phi}_X + \bar{\phi}_T) < 1$; solutions outside these limits imply steady state cannot be reached under the assumed constraints. The equations can be solved uniquely if $\overline{[\mathrm{MgO}]}_o$ and $\bar{D}_{Mg}$ are specified.

The average MgO content of parental OFB (that is, $\overline{[\mathrm{MgO}]}_o$) has remained controversial for decades, because the MgO content reflects the mantle potential temperature[22]. We circumvent the controversy by noting that $\overline{[\mathrm{MgO}]}_o$ can be interpreted as the composition near the start of low-pressure magma-chamber evolution, and not necessarily as the 'primary' composition extracted from the mantle. Figure 4a shows the histogram [MgO] for the 9,050 OFB glasses in the Smithsonian data file[12]. Only one glass has more than 10 wt% MgO (10.07 wt%). We therefore take $\overline{[\mathrm{MgO}]}_o$ to be 10.0 wt% or thereabouts. This is approximately the minimum required by several phase equilibrium constraints—in particular, the contraction of the primary phase volume of olivine with increasing pressure[23]. The value of $\bar{D}_{\mathrm{MgO}}$ along the ol+pl+cpx cotectic as modelled in Fig. 1 increases from 1.3 at 8.3 wt% MgO to 1.8 at 5.5 wt% MgO, but the replenishing magma would move liquid compositions off the cotectic into the ol±pl primary phase volumes, giving initially higher $\bar{D}_{\mathrm{Mg}}$. We make the simplifying assumption that $\bar{D}_{\mathrm{Mg}}$ is constant, with a value determined by the fractions of phases crystallizing.

Solutions satisfying all constraints (allowing for uncertainties in the literature values of $\bar{D}_{\mathrm{M}}$) are obtained with constant $\bar{D}_{\mathrm{MgO}}$ at values near 2, which we optimized to 1.9, while increasing $\overline{[\mathrm{MgO}]}_o$ to 10.4 wt% so that steady-state behaviour begins at $\overline{[\mathrm{MgO}]}_T = 8.6$ wt%, which is similar to the parental composition given in ref. 4. Steady-state behaviour terminates at $\overline{[\mathrm{MgO}]}_T = 5.6$ wt% (with $\bar{\phi}_T = 0$; Fig. 4b), consistent with the distribution of [MgO] in OFB glasses (Fig. 4a). The implication of our model is that liquids with [MgO] >8.6 wt% are rare because these compositions either bypass magma chambers or come from juvenile magma chambers that have yet to reach steady state, possibly contributing to the higher variability of P_2O_5 and K_2O in the most Mg-rich basalts (Fig. 2c). Liquids erupted with [MgO] <5.6 wt% come from dying chambers that are not being recharged.

Calculated $\log([M]_T/[M]_o)$ for various values of $\bar{D}_{\mathrm{M}}$ are shown as a function of [MgO] in Fig. 4c. The trends are linear, as required empirically (Fig. 1d–f). We use 19 trace elements selected because they have relatively well constrained partition coefficients (ten REEs; four other RLEs, namely Sc, Ti, Sr and Zr; plus the five SVEs, namely Li, Na, Mn, Co and Ni) to calculate the best match between $\bar{D}_{\mathrm{M}}$ obtained from slopes and the crystal/melt partition coefficients from the literature; we do this by least-squares optimization of $\bar{f}^{\mathrm{pl}}, \bar{f}^{\mathrm{cpx}}$ and $\bar{f}^{\mathrm{ol}}$, with the constraints $\bar{f}^{\mathrm{pl}} + \bar{f}^{\mathrm{cpx}} + \bar{f}^{\mathrm{ol}} = 1$ and $\bar{D}_{\mathrm{Mg}} = 1.9$. We find $\bar{f}^{\mathrm{ol}} = 0.18$, $\bar{f}^{\mathrm{pl}} = 0.41$ and $\bar{f}^{\mathrm{cpx}} = 0.42$ (details in Methods). With an uncertainty of 20% in crystal/melt partition coefficients, the mean square weighted deviation (MSWD) for the fit is 1.4, indicating good agreement between model and data (Fig. 5). For the enriched Macquarie Island basalts, slopes are higher for the cpx-hosted trace elements such as the heavy REE (Fig. 2e), but lower for plagioclase-hosted Sr, Eu and Na, signifying that they have evolved by crystallization with much higher cpx/pl. The least-squares optimization of the model for these basalts gives $\bar{f}^{\mathrm{ol}} = 0.06$, $\bar{f}^{\mathrm{pl}} = 0.09$, $\bar{f}^{\mathrm{cpx}} = 0.85$ and $\bar{D}_{\mathrm{Mg}} = 2.2$, assuming $\overline{[\mathrm{MgO}]}_o = 10.4$ wt%.

The global solution infers a relationship between $\bar{\phi}_X$ and $\bar{\phi}_T$ in the ensemble of OFB magma chambers (Fig. 4b). These two variables control the average MgO content of the magma chamber (that is, $\overline{[\mathrm{MgO}]}_T$), which varies nearly linearly with the average fraction of magma retained after each cycle (that is, $1 - \bar{\phi}_X - \bar{\phi}_T$). Either there is some process by which $\bar{\phi}_X$ determines $\bar{\phi}_T$ on average, or vice versa. Whether this is plausible is a matter for geophysical fluid dynamics and is raised here as a hypothesis. It is easier to visualize physical reasons why magma chambers might erupt a constant fraction of their contents (that is, $\bar{\phi}_T$ constant), but such models do not produce linear trends of log[M] versus [MgO] under our simplifying assumption of constant partition coefficients. However, an elaborated model—with values of $\bar{D}_{\mathrm{M}}$ and $\bar{D}_{\mathrm{Mg}}$ allowed to vary with [MgO] and decreasing temperature, or with related indices of evolution (such as the anorthite content of plagioclase), as indicated by experimental partitioning studies (see, for example, refs 24–26)—may provide satisfactory solutions. The model does not depend on the sizes of the magma

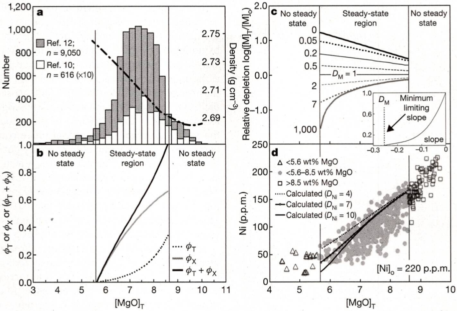


Figure 4 | The global ensemble of replenished and tapped magma chambers. **a**, Histogram of [MgO] in OFB glasses from the Smithsonian abyssal volcanic glass data file[12] (shaded bars), with the subset of samples analysed for trace elements in ref. 2 shown by the white bars. Also shown (right-hand *y* axis, dot-dashed line) are the densities of liquids for the model given in Fig. 1, calculated with PETROLOG[42,43]. The frequency distribution of [MgO] does not correspond to the density minimum. **b**, Solution to the steady-state magma chamber model with constant $\overline{D}_{\mathrm{Mg}} = 1.9$ and $\overline{[\mathrm{MgO}]}_{\mathrm{o}} = 10.4$ wt%. $\bar{\phi}_{\mathrm{T}}$ is the average fraction of the average magma chamber tapped per cycle, and $\bar{\phi}_{\mathrm{X}}$ is the average fraction crystallized. Steady state is achieved only within the limits $0 < (\bar{\phi}_{\mathrm{T}} + \bar{\phi}_{\mathrm{X}}) < 1$. The region of steady state coincides with the distribution of [MgO], which explains why OFB with >8.6 wt% MgO or <5.6 wt% MgO are uncommon. **c**, $\log([\mathrm{M}]_{\mathrm{T}}/[\mathrm{M}]_{\mathrm{o}})$ against [MgO] for $\overline{D}_{\mathrm{Mg}} = 1.9$ and $\overline{[\mathrm{MgO}]}_{\mathrm{o}} = 10.4$ wt%. Slopes d(log[M])/d[MgO] are calculated to be constant, except for the hypothetical case of a very highly compatible element. Inset, dependence of D_{M} on the slopes. A distinctive feature of the model is that slopes are most sensitive to D_{M} for elements with $0 < D_{\mathrm{M}} < 0.1$ (that is, highly incompatible), but with the slopes nevertheless varying nearly linearly with D_{M} in this region until a cut-off is reached at the minimum limiting slope. **d**, Decrease of compatible element [Ni] with [MgO] in OFB glasses[10] compared to that calculated using our model, with $[\mathrm{Ni}]_{\mathrm{o}} = 220$ p.p.m., $\overline{D}_{\mathrm{Mg}} = 1.9$ and $\overline{[\mathrm{MgO}]}_{\mathrm{o}} = 10.4$ wt%. A parental liquid with $[\mathrm{Ni}]_{\mathrm{o}} = 220$ would be in equilibrium with mantle olivine with 3,300 p.p.m. Ni if $D_{\mathrm{Ni}}^{\mathrm{ol/melt}} = 15$ (see Supplementary Information), precluding significant olivine-only fractionation.

chambers, which may be too small in places to be imaged by geophysical means. The ratio $\bar{\phi}_{\mathrm{T}}/\bar{\phi}_{\mathrm{X}}$ changes from 0.5 at 8.6 wt% MgO to 0.06 at 5.6 wt% MgO, leading to a prediction that the ratio of basalt+dolerite to cumulus gabbro in oceanic crust might vary similarly, correlating with the MgO content of the basalts.

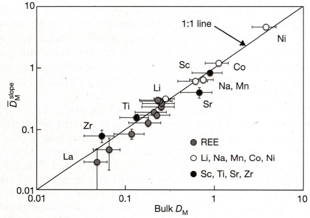


Figure 5 | Testing the model. Shown is a comparison of bulk partition coefficients $\overline{D}_{\mathrm{M}}^{\mathrm{slope}}$ calculated from the observed slopes d(log[M])/d[MgO] assuming the model values of $\bar{\phi}_{\mathrm{T}}$ and $\bar{\phi}_{\mathrm{X}}$ (see Fig. 4b), with bulk partition coefficients calculated from experimentally or empirically determined mineral/melt partition coefficients from the literature, at the model's best-fit values of $\bar{f}^{\mathrm{cpx}} = 0.42$, with $\bar{f}^{\mathrm{pl}} = 0.41$ and $\bar{f}^{\mathrm{ol}} = 0.18$. Error bars are ±2 s.d. Good agreement is obtained for all elements including the compatible element Ni, with a reduced χ^2 (MSWD) of 1.4. See Methods and Supplementary Table 1 for the partition coefficients used in the modelling.

Incompatible elements are fractionated from each other in proportion to the absolute difference in their bulk partition coefficients (that is, as $\overline{D}_{\mathrm{M1}} - \overline{D}_{\mathrm{M2}}$). Such fractionation is much greater than can be achieved by simple fractional crystallization, and can change the ratios of all but the most highly incompatible elements. Thus the model can explain the observed differences between slopes of K, U and La, which have $\overline{D}_{\mathrm{M}} \approx 0.01$–$0.05$, and the difference between these slopes and those of the even more incompatible elements (Ba, Th, Nb, and so on). Only for these latter elements, which share the limiting minimum slope, does the model predict that their ratios should remain constant. Nb/Th and Rb/Ba would be examples.

The influence of plagioclase can be used to distinguish between crustal and mantle processes, because plagioclase is not stable in the pressure interval over which partial melting in the mantle takes place, but is a major phase in low-pressure crystallization (see, for example, refs 5, 6, 27). For the majority of incompatible trace elements, clinopyroxene is the main host phase during both mantle melting and crustal evolution, which accounts for the good correlation between intercepts and slopes (Fig. 3a). Sc is one exception, because its compatibility in orthopyroxene allows a significant proportion to be retained in the mantle residue of partial melting[24]. Anomalous behaviour is also shown by Sr and Eu. Because of their compatibility in plagioclase (for example, $D_{\mathrm{Sr}}^{\mathrm{pl/melt}} \approx 1.6$; ref. 28), Sr and Eu plot shifted towards the right in Fig. 3a relative to their normal geochemical compatibility among the REEs (Sr is considered to be similar in incompatibility to Nd during mantle melting, and Eu should fall between Sm and Gd). So does Pb, which is also more compatible in plagioclase than clinopyroxene ($D_{\mathrm{Pb}}^{\mathrm{pl/melt}} \approx 0.1$; ref. 29), although fractionation of Pb by immiscible sulphide matte (responsible for the compatibility of Cu and Ag in OFB evolution[30]) may additionally contribute. The slope of Pb is similar to that of Ce during low-pressure evolution, but its depletion in the mantle source, indicated by its normalized intercept value, is closer to that of U, providing a reconciliation of the 'third Pb paradox'[31]. The variability of Sr stands out as larger than that of other elements with similar slopes (Fig. 3b), but is consistent with its incompatibility (similar to Nd) during mantle processes.

Another element much more compatible in plagioclase than clinopyroxene, especially at low pressures, is Na. As the Macquarie Island basalts show, variations in Na contents in OFB at a given [MgO] may reflect the variations in the ratio of pl/cpx that have fractionated.

Variability of trace elements in OFB

Much of the interest in OFB geochemistry comes from the expectation that part of the variations in their composition may be due to plate-tectonic variables, including the thermal structure of the mantle. For example, variations in Na, Fe and Ca/Al correlate with plate spreading rate and perhaps axial depth of mid-ocean ridges[2,3,32–34]. A key question is what proportion of the observed variations arises from low-pressure evolution, and what proportion is due to mantle processes, or source composition. The entity that we call variability addresses this problem. Note that our definition of variability (see Fig. 1 and Supplementary Information) differs from that used previously[14,15], because we calculate it relative to the global trends, assuming probability density functions of log-normal distributions with the mean depending linearly on [MgO].

Although the very highly incompatible elements (Ba, Nb, Th and so on) are not fractionated from each other by the magma-chamber processes, they show the highest variabilities (Fig. 3), which are therefore ascribed to variability of the parental magmas (that is, in $[\mathrm{M}]_\mathrm{o}$). Potentially this may derive from variations in extent of melting, the mechanism of melt extraction (for example, residual porosity), or be inherited from the mantle source. The distributions of radiogenic isotopic ratios in OFBs allow us to monitor the heterogeneity of source compositions. There is an imperfect correlation ($R \approx -0.5$) globally between La/Sm and $^{143}\mathrm{Nd}/^{144}\mathrm{Nd}$, which reflects time-averaged Sm/Nd (see figure 14 in ref. 35). However, the variations of $^{143}\mathrm{Nd}/^{144}\mathrm{Nd}$ in modern basalts are a function of both the extents of the fractionation of parent Sm from daughter Nd in the source, and the times at which the fractionations happened. Subsequent fractionation of La/Sm (and hence by implication, Sm/Nd) either by melting processes, or, as shown to be possible here, during low-pressure differentiation, will further blur the correlation. Although source effects are quite important, isotope/trace-element correlations cannot quantify them very well at this point.

Variability will accrue from the many ways that the magma chambers depart from both steady state and the global average. The basic principle is that a magma will be more enriched than the average at the same [MgO] by inheriting a higher proportion of enriched magma from previous cycles, and vice versa. Magma may also be erupted out of sequence—for example, from a chamber undergoing fractional crystallization between replenishments. Basalts erupted in one locality over a few years, such as those from well-studied segments of the East Pacific Rise[36,37], probably record such departures. Compositions from these suites plot along short chords of log[M] versus [MgO] arranged '*en échelon*' at low angles across the global trends, as expected for simple fractional crystallization. We envisage the global trends as the averages of many such chords.

The number of cycles needed to reach steady state increases with the incompatibility of the element[21]. Although this may account for some of the increased variability of the most incompatible elements, mixing in the magma chambers acts to homogenize variability in parental magmas. Melt inclusions trapped in growing crystals may preserve incompletely homogenized aliquots of replenishing magma, which explains why they often record greater variability than bulk magmas[33]. Detailed interpretations of the chemical variability in short ridge segments has highlighted the complexity in individual magma chambers undergoing replenishment[38,39]. But when taken over the global ensemble of magma chambers, this complexity averages out to give the systematic global trends shown in Figs 2 and 3.

Implications for OFB petrogenesis

The 'cotectic illusion' (Fig. 1) is caused by crystallization in replenished magma chambers forcing liquids onto the cotectic, on which magma compositions appear "perched"[19,40]. Erupted major-element compositions are indistinguishable from those produced by fractional crystallization along the cotectic, but their trace element systematics differ greatly. A series of compositions plotting along a cotectic may not be a 'liquid line of descent' related by simple fractional crystallization, but the locus of different fractionation paths, each ending on the cotectic. The assumption that such liquid compositions reflect the depth of the magma chamber in which they evolved remains valid[5,6], but the many inferences drawn from OFB chemistry that have assumed the paradigm of simple fractional crystallization in order to see beyond their low-pressure evolution require re-evaluation.

Simple fractional crystallization is inefficient at changing the ratios of incompatible elements. There has therefore been a presumption that ratios of even quite moderately incompatible trace elements like the mid to heavy REEs in OFB are inherited from their parental compositions. On the contrary, replenished magma chambers are effective at changing incompatible-element ratios. Only ratios among elements with $\overline{D}_\mathrm{M} < 0.01$ (for example, Ba, Nb, Ta and Th and the SVEs Rb, Cs and W) are expected to escape change during low-pressure evolution.

Uranium-series disequilibria provide an example of how the replenished magma chamber model necessitates re-evaluation of petrogenetic interpretations. Uranium-series disequilibria measure Th/U fractionations over short timescales ($<10^4$ years), so cannot reflect source heterogeneity, but have been ascribed to in-growth during partial melting[41]. Fractional crystallization has been assumed to be inconsequential. However, using our solution to the average global pattern, the fraction of a perfectly incompatible element ($D_\mathrm{M} = 0$) inherited from previous cycles of the global-average magma chamber is 81% at 8 wt% MgO and 93% at 6 wt% MgO, raising the possibility that U-series secular disequilibria may reflect in-growth in the magma chamber. Among the global ensemble of magma chambers there will be those whose characteristics will produce even more extreme effects.

METHODS SUMMARY

Trace-element contents of OFB glasses were fitted to the equation:

$$\log[\mathrm{M}] = \log[\mathrm{M}]_\mathrm{o} + \mathrm{slope}([\mathrm{MgO}] - 10)$$

by minimizing χ^2:

$$\chi^2 = \frac{1}{\sigma(\log[M])^2}\sum\left(\log[\mathrm{M}] - \log[\mathrm{M}]_\mathrm{o} - \mathrm{slope}(10 - [\mathrm{MgO}])\right)^2$$

assuming that $\sigma(\log[M])$ is constant. If χ^2/n is put equal to 1, its most probable value, then:

$$\sigma(\log[\mathrm{M}]) = \left(\frac{\sum(\log[\mathrm{M}] - \log[\mathrm{M}]_\mathrm{calc})^2}{n}\right)^{1/2}$$

where n is the number of data and $\log[\mathrm{M}]_\mathrm{calc} = \log[\mathrm{M}]_\mathrm{o} - \mathrm{slope}(10 - [\mathrm{MgO}])$. We take $\sigma(\log[\mathrm{M}])$ as the definition of variability.

To facilitate comparison, intercept values for RLEs at 10 wt% MgO are plotted normalized to CI chondrites using data in ref. 17. The intercept value of Pb was likewise normalized by dividing by the bulk silicate Earth value of 0.185 p.p.m. obtained from $^{238}\mathrm{U}/^{204}\mathrm{Pb} = 8.5$, and multiplying by the RLE enrichment factor of 2.8 (ref. 17).

The evaluation of the steady-state magma-chamber model proceeds by first calculating, as a function of the average tapped MgO content, $\overline{[\mathrm{MgO}]}_\mathrm{T}$, the parameters $\bar{\phi}_\mathrm{T}$ and $\bar{\phi}_\mathrm{X}$, the global averages of the fractions tapped and crystallized, where the averages are to be thought of as taken over the global ensemble of magma chambers. This initial stage of the calculation uses the observation that the slopes of the logarithms of the concentrations versus MgO for all the very highly incompatible elements (that is, $d(\log[M_H])/d[MgO]$) share a constant value of −0.26 (Fig. 3). These calculated parameters were used to calculate the slopes for bulk partition coefficients greater than zero (see Fig. 4c). These relationships were then used to convert the observed slopes for 19 elements to bulk partition coefficients, $\bar{D}_M^{slope}$. Finally, values of $\bar{D}_M^{slope}$ were matched to bulk partition coefficients calculated from crystal/melt partition coefficients, obtained from experiment or natural phenocryst/matrix pairs, as shown in Fig. 5. The model has three parameters, the global average fractions of plagioclase, clinopyroxene and olivine crystallizing ($\bar{f}^{pl}$, $\bar{f}^{cpx}$ and $\bar{f}^{ol}$, respectively), only one of which is independently variable. A fuller description is given in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 7 March; accepted 15 October 2012.

1. Klein, E. M. in *Treatise on Geochemistry* Vol. 3, *The Crust* (ed. Rudnick, R. L.) 433–463 (Pergamon, 2003).
2. Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92** (B8), 8089–8115 (1987).
3. Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Morgan, J. P., Blackman, D. K. & Sinton, J. M.) 183–280 (Geophysical Monograph Series Vol. 71, AGU, 1992).
4. McKenzie, D. & Bickle, M. J. The volume and composition of melt generated by extension of the lithosphere. *J. Petrol.* **29,** 625–679 (1988).
5. Herzberg, C. Partial crystallization of mid-ocean ridge basalts in the crust and mantle. *J. Petrol.* **45,** 2389–2405 (2004).
6. Michael, P. J. & Cornell, W. C. Influence of spreading rate and magma supply on crystallization and assimilation beneath mid-ocean ridges: evidence from chlorine and major element chemistry of mid-ocean ridge basalts. *J. Geophys. Res.* **103,** 18325–18356 (1998).
7. O'Hara, M. J. Are ocean floor basalts primary magmas? *Nature* **220,** 683–686 (1968).
8. Dungan, M. A. & Rhodes, J. M. Residual glasses and melt inclusions in basalts from DSDP Legs 45 and 46: evidence for magma mixing. *Contrib. Mineral. Petrol.* **67,** 417–431 (1978).
9. White, W. M. & Bryan, W. B. Sr-isotope, K, Rb, Cs, Sr, Ba, and rare-earth geochemistry of basalts from the FAMOUS area. *Geol. Soc. Am. Bull.* **88,** 571–576 (1977).
10. Jenner, F. E. & O'Neill, H. St C. Analysis of 60 elements in 616 ocean floor basaltic glasses. *Geochem. Geophys. Geosyst.* **13,** Q02005, http://dx.doi.org/10.1029/2011GC004009 (2012).
11. Arevalo, R. & McDonough, W. F. Chemical variations and regional diversity observed in MORB. *Chem. Geol.* **271,** 70–85 (2010).
12. Melson, W. G., O'Hearn, T. & Jarosewich, E. A data brief on the Smithsonian abyssal volcanic glass data file. *Geochem. Geophys. Geosyst.* **3,** 1–11 (2002).
13. Wood, B. J. & Blundy, J. D. in *Treatise on Geochemistry* Vol. 2, *The Mantle and Core* (ed. Carlson, R. W.) 395–424 (Elsevier, 2003).
14. Hofmann, A. W. Chemical differentiation of the Earth: the relationship between mantle, continental crust, and oceanic crust. *Earth Planet. Sci. Lett.* **90,** 297–314 (1988).
15. Schiano, P., Allègre, C. J., Dupré, B., Lewin, E. & Joron, J.-L. Variability of trace elements in basaltic suites. *Earth Planet. Sci. Lett.* **119,** 37–51 (1993).
16. Kamenetsky, V. S. & Eggins, S. M. Systematics of metals, metalloids, and volatiles in MORB melts: effects of partial melting, crystal fractionation and degassing (a case study of Macquarie Island glasses). *Chem. Geol.* **302–303,** 76–86 (2012).
17. Palme, H. & O'Neill, H. St C. in *Treatise on Geochemistry* Vol. 2, *The Mantle and Core* (ed. Carlson, R. W.) 1–38 (Elsevier, 2003).
18. Kendrick, M. A., Kamenetsky, V. S., Phillips, D. & Honda, M. The behaviour of halogens (Cl, Br, I) in enriched mid-ocean ridge basalts and their distribution on Earth. *Geochim. Cosmochim. Acta* **81,** 82–93 (2012).
19. O'Hara, M. J. Geochemical evolution during fractional crystallisation of a periodically refilled magma chamber. *Nature* **266,** 503–507 (1977).
20. O'Hara, M. J. & Mathews, R. E. Geochemical evolution in an advancing, periodically replenished, periodically tapped, continuously fractionated magma chamber. *J. Geol. Soc. Lond.* **138,** 237–277 (1981).
21. Albarède, F. Regime and trace-element evolution of open magma chambers. *Nature* **318,** 356–358 (1985).
22. Falloon, T. J., Danyushevsky, L. V., Ariskin, A., Green, D. H. & Ford, C. E. The application of olivine geothermometry to infer crystallisation temperatures of parental liquids: implications for the temperatures of MORB magmas. *Chem. Geol.* **241,** 207–233 (2007).
23. Hess, P. C. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Morgan, J. P., Blackman, D. K. & Sinton, J. M.) 67–102 (Geophysical Monograph Series Vol. 71, AGU, 1992).
24. Beattie, P., Ford, C. & Russell, D. Partition coefficients for olivine-melt and orthopyroxene-melt systems. *Contrib. Mineral. Petrol.* **109,** 212–224 (1991).
25. Bindeman, I. N., Davis, A. M. & Drake, M. J. Ion microprobe study of plagioclase-basalt partition experiments at natural concentration levels of trace elements. *Geochim. Cosmochim. Acta* **62,** 1175–1193 (1998).
26. Wood, B. J. & Blundy, J. D. A predictive model for rare earth element partitioning between clinopyroxene and anhydrous silicate melt. *Contrib. Mineral. Petrol.* **129,** 166–181 (1997).
27. Falloon, T. J., Green, D. H., Danyushevsky, L. V. & McNeill, A. W. The composition of near-solidus partial melts of fertile peridotite at 1 and 1·5 GPa: implications for the petrogenesis of MORB. *J. Petrol.* **49,** 591–613 (2008).
28. Aigner-Torres, M., Blundy, J., Ulmer, P. & Pettke, T. Laser ablation ICPMS study of trace element partitioning between plagioclase and basaltic melts: an experimental approach. *Contrib. Mineral. Petrol.* **153,** 647–667 (2007).
29. Leeman, W. P. Partitioning of Pb between volcanic glass and coexisting sanidine and plagioclase feldspars. *Geochim. Cosmochim. Acta* **43,** 171–175 (1979).
30. Jenner, F. E., O'Neill, H. St C., Arculus, R. J. & Mavrogenes, J. A. The magnetite crisis in the evolution of arc-related magmas and the initial concentration of Au, Ag, and Cu. *J. Petrol.* **51,** 2445–2464 (2010).
31. Hart, S. R. & Gaetani, G. A. Mantle Pb paradoxes: the sulfide solution. *Contrib. Mineral. Petrol.* **152,** 295–308 (2006).
32. Dick, H. J. B. & Bullen, T. Chromian spinel as a petrogenetic indicator in abyssal and alpine-type peridotites and spatially associated lavas. *Contrib. Mineral. Petrol.* **86,** 54–76 (1984).
33. Rubin, K. H. & Sinton, J. M. Inferences on mid-ocean ridge thermal and magmatic structure from MORB compositions. *Earth Planet. Sci. Lett.* **260,** 257–276 (2007).
34. Shen, Y. & Forsyth, D. W. Geochemical constraints on initial and final depths of melting beneath mid-ocean ridges. *J. Geophys. Res.* **100,** 2211–2237 (1995).
35. Hofmann, A. W. in *Treatise on Geochemistry* Vol. 2, *The Mantle and Core* (ed. Carlson, R. W.) 61–101 (Elsevier, 2003).
36. Goss, A. R. *et al.* Geochemistry of lavas from the 2005–2006 eruption at the East Pacific Rise, 9°46′N–9°56′N: implications for ridge crest plumbing and decadal changes in magma chamber compositions. *Geochem. Geophys. Geosyst.* **11,** Q05T09, http://dx.doi.org/10.1029/2009GC002977 (2010).
37. Sims, K. W. W. *et al.* Chemical and isotopic constraints on the generation and transport of magma beneath the East Pacific Rise. *Geochim. Cosmochim. Acta* **66,** 3481–3504 (2002).
38. Rannou, E., Caroff, M. & Cordier, C. A geochemical approach to model periodically replenished magma chambers: does oscillatory supply account for the magmatic evolution of EPR 17–19°S? *Geochim. Cosmochim. Acta* **70,** 4783–4796 (2006).
39. Cordier, C. Caroff, M. & Rannou, E. Timescale of open-reservoir evolution beneath the south Cleft segment, Juan de Fuca ridge. *Mineral. Petrol.* **104,** 1–14 (2012).
40. Walker, D., Shibata, T. & DeLong, S. Abyssal tholeiites from the Oceanographer fracture zone. *Contrib. Mineral. Petrol.* **70,** 111–125 (1979).
41. Elliott, T. & Spiegelman, M. in *Treatise on Geochemistry* Vol. 2, *The Mantle and Core* (ed. Carlson, R. W.) 465–510 (Elsevier, 2003).
42. Danyushevsky, L. V. & Plechov, P. Petrolog3: integrated software for modeling crystallization processes. *Geochem. Geophys. Geosyst.* **12,** Q07021, http://dx.doi.org/10.1029/2011GC003516 (2011).
43. Danyushevsky, L. V. The effect of small amounts of H_2O on crystallisation of mid-ocean ridge and backarc basin magmas. *J. Volcanol. Geotherm. Res.* **110,** 265–280 (2001).
44. Cottrell, E. & Kelley, K. A. The oxidation state of Fe in MORB glasses and the oxygen fugacity of the upper mantle. *Earth Planet. Sci. Lett.* **305,** 270–282 (2011).
45. Jenner, F. E. & O'Neill, H. St C. Major and trace analysis of basaltic glasses by laser-ablation ICP-MS. *Geochem. Geophys. Geosyst.* **13,** Q03003, http://dx.doi.org/10.1029/2011GC003890 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by the Australian National University. I. Campbell, R. Arculus, S. Turner, R. Carlson and E. Hauri are thanked for comments on earlier versions of this manuscript, and the final presentation has benefited from reviews by A. Hofmann and W. McDonough.

Author Contributions Both authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.St C.O'N. (hugh.oneill@anu.edu.au).

ARTICLE

doi:10.1038/nature11650

Analysis of the bread wheat genome using whole-genome shotgun sequencing

Rachel Brenchley[1], Manuel Spannagl[2], Matthias Pfeifer[2], Gary L. A. Barker[3], Rosalinda D'Amore[1], Alexandra M. Allen[3], Neil McKenzie[4], Melissa Kramer[5], Arnaud Kerhornou[6], Dan Bolser[6], Suzanne Kay[1], Darren Waite[4], Martin Trick[4], Ian Bancroft[4], Yong Gu[7], Naxin Huo[7], Ming-Cheng Luo[8], Sunish Sehgal[9], Bikram Gill[9], Sharyar Kianian[10], Olin Anderson[7], Paul Kersey[6], Jan Dvorak[8], W. Richard McCombie[5], Anthony Hall[1], Klaus F. X. Mayer[2], Keith J. Edwards[3], Michael W. Bevan[4] & Neil Hall[1]

Bread wheat (*Triticum aestivum*) is a globally important crop, accounting for 20 per cent of the calories consumed by humans. Major efforts are underway worldwide to increase wheat production by extending genetic diversity and analysing key traits, and genomic resources can accelerate progress. But so far the very large size and polyploid complexity of the bread wheat genome have been substantial barriers to genome analysis. Here we report the sequencing of its large, 17-gigabase-pair, hexaploid genome using 454 pyrosequencing, and comparison of this with the sequences of diploid ancestral and progenitor genomes. We identified between 94,000 and 96,000 genes, and assigned two-thirds to the three component genomes (A, B and D) of hexaploid wheat. High-resolution synteny maps identified many small disruptions to conserved gene order. We show that the hexaploid genome is highly dynamic, with significant loss of gene family members on polyploidization and domestication, and an abundance of gene fragments. Several classes of genes involved in energy harvesting, metabolism and growth are among expanded gene families that could be associated with crop productivity. Our analyses, coupled with the identification of extensive genetic variation, provide a resource for accelerating gene discovery and improving this major crop.

With a global output of 681 million tonnes in 2011[1], bread wheat accounts for 20% of the calories consumed by humans[2] and is an important source of protein, vitamins and minerals. It originated from hybridization between cultivated tetraploid emmer wheat (AABB, *Triticum dicoccoides*) and diploid goat grass (DD, *Aegilops tauschii*) approximately 8,000 years ago[3]. Bread wheat cultivation and domestication has been directly associated with the spread of agriculture and settled societies, and it is now one of the most widely cultivated crops owing to its high yields and nutritional and processing qualities. The three diploid progenitor genomes, AA from *Triticum urartu*, BB from a species that is unknown but which may be of the section *Sitopsis* (to which *Aegilops speltoides* belongs), and DD from *Ae. tauschii*, radiated from a common Triticeae ancestor between 2.5 and 4.5 million years ago, and AABB tetraploids arose less than 0.5 million years ago[4,5]. Nucleotide diversity in the AABB and DD genomes is substantially reduced compared with ancestral populations, indicating a major diversity bottleneck on the transition to cultivated lines[6].

Grass genomes show extensive long-range conservation of gene order[7–9]. Nevertheless, they are highly dynamic owing to the activities of repeats that contribute to tremendous variation in genome size[10], changes in local gene order and pseudogene formation, particularly in larger genomes such as those of maize[11] and wheat[12]. From analysis of BAC contigs on chromosome 3B, the 17-gigabase-pair (Gb) genome was estimated to be composed of approximately 80% repeats, primarily retroelements, with a gene density of between 1 per 87 kilobase pairs and 1 per 184 kilobase pairs[13]. Despite both the substantial knowledge gained of the wheat genome from these studies and the central importance of the wheat crop, a comprehensive genome-wide analysis of gene content has yet to be conducted owing to its large size, repeat content and polyploid complexity.

We have analysed a low-coverage, long-read (454) shotgun sequence of the hexaploid wheat genome using gene sequences from diverse grasses. From this, we created assemblies of wheat genes in an orthologous gene family framework, used diploid wheat relatives to classify homeologous relationships, and defined a genome-wide catalogue of single nucleotide polymorphisms (SNPs) in the A, B and D genomes. These analyses provide a foundation for genetic and genomic analysis of this key crop.

Sequence analysis

The wheat variety Chinese Spring (CS42) was selected for sequencing because of its wide use in genome studies[14,15]. Purified nuclear DNA was sequenced using Roche 454 pyrosequencing technology (GS FLX Titanium and GS FLX+ platforms) to generate 85 Gb of sequence (220 million reads), corresponding to approximately five-fold coverage on the basis of an estimated genome size of 17 Gb. Supplementary Table 1 shows that 79% of the reads had matches to the Triticeae Repeat Sequence Database, and most hit retrotransposons, consistent with previous studies[13]. To identify A-, B- and D-genome-derived gene assemblies in the hexaploid sequences, we used Illumina sequence assemblies of *Triticum monococcum*, related to the A-genome donor, *Ae. speltoides* complementary DNA (cDNA) assemblies and 454 sequences from the D-genome donor *Ae. tauschii*, respectively. The SOLiD platform was used to generate additional sequence of CS42 and three commercial wheat varieties to increase the accuracy of homeologous SNP identification. Data sets are summarized in Table 1 and Supplementary Table 2, and SNP

[1]Centre for Genome Research, University of Liverpool, Liverpool L69 7ZB, UK. [2]MIPS/IBIS, Helmholtz-Zentrum München, 85764 Neuherberg, Germany. [3]School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK. [4]John Innes Centre, Norwich NR4 7UH, UK. [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [6]European Bioinformatics Institute, Hinxton CB10 1SD, UK. [7]USDA Western Regional Laboratory, Albany, California 94710, USA. [8]Department of Plant Sciences, University of California, Davis, California 95616, USA. [9]Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506, USA. [10]Department of Plant Sciences, North Dakota State University, Fargo, North Dakota 58018-6050, USA.

Table 1 | Sequence sources used for analysis

Genome	Platform	Size of data set	Reference
T. aestivum (CS42) genomic DNA	454 GS FLX Titanium/454 GS FLX+	85 Gb	EBI study: ERP000319
T. aestivum (CS42) genomic DNA from sorted chromosomes 1A, 1B and 1D	454 GS FLX Titanium	1A: 287 Mb 1B: 392 Mb 1D: 375 Mb	Ref. 12
T. aestivum (CS42, Avalon, Rialto, Savannah) genomic DNA	SOLiD 3/SOLiD 4	15.2 billion reads	EBI study: ERP001493
T. aestivum (CS42) cDNA	454 GS FLX Titanium/454 GS FLX+	1.6 Gb	EBI study: ERP001415
T. monococcum genomic DNA	Illumina GAIIx/HiSeq	A/B/D sequences: 3.7 Gb A/B/D SNPs: 401 Gb	NCBI archive: SRP004490.3
Ae. speltoides cDNA	Pre-assembled data	151 Mb	M. Trick and I. Bancroft, unpublished observations
Ae. tauschii genomic DNA	454 GS FLX Titanium	12.8 Gb	M.-C.L. *et al.*, submitted
Ae. tauschii genomic DNA	SOLiD 4	80–100-fold coverage	J. Dvorak, unpublished observations

EBI, European Bioinformatics Institute; NCBI, US National Center for Biotechnology Information.

identification methods are described in Supplementary Information, section 5.2.

Sequence assembly

An orthologous group assembly (Supplementary Table 3) was created by clustering 454 reads by sequence similarity to orthologous grass gene sequences, and separate assembly of the clusters at high stringency using Newbler (Supplementary Information, section 2). The orthologous genes were derived from rice[16], sorghum[8], *Brachypodium*[9] and barley full-length cDNAs by OrthoMCL[17] clustering. This generated 20,496 orthologous groups (Supplementary Table 4 and Supplementary Fig. 1). The gene model with highest similarity to wheat (termed the orthologous group representative (OGR)) was selected from each orthologous group by stringent BLASTX comparison to a low-copy-number genome assembly (LCG) made by filtering out repetitive sequences and assembling the remaining low-copy-number sequences *de novo* (Supplementary Table 3). The assemblies are described in Table 2. Nearly 90% of the metabolic genes in *Arabidopsis* matched OGRs, and the 20,051 OGRs matched 92% of publicly available wheat full-length cDNAs[18] and 78.7% of the harvEST set of wheat cDNA assemblies (Supplementary Fig. 2), indicating that they represent nearly all wheat genes.

We optimized parameters for wheat gene assembly using MetaSim[19] to generate simulated fivefold 454 reads from the allotetraploid maize genome and from a triplicated rice gene set, with the introduction of sequence variation (Supplementary Information, section 2.7). Similar degrees of coverage over the OGRs were seen for the simulated data sets and wheat 454 reads (Fig. 1a). Rice reads followed the same depth distribution as the wheat reads (Fig. 1b), suggesting that they are a reasonable representation of hexaploid sequences. Maize reads covered their OGRs to a median depth of approximately five, consistent with fivefold coverage.

Simulated maize and triplicated rice 454 reads were used to optimize assembly parameters. Assembly at 99% minimum sequence identity (m.i.) using 40-bp overlap length predicted gene family sizes most accurately (Supplementary Figs 3–6). Wheat 454 reads were pre-processed (Supplementary Table 5) and assembled using 99% m.i. (Supplementary Tables 6 and 7) to create the orthologous group assembly. Figure 1b shows that the depth of coverage of the orthologous group assembly followed a similar pattern to maize, consistent with multiple gene copies. In contrast, the low depth coverage by the LCG assembly suggested that gene family numbers were collapsed. The number of wheat assemblies for each OGR was calculated to determine gene copy numbers (Supplementary Table 7). Figure 1c shows that most OGRs had between one and five distinctive wheat gene assemblies, with a peak of two genes.

The A, B and *Ae. tauschii* (D) genomes[13,20,21] have been estimated to contain approximately 28,000, 38,000 and 36,000 genes, respectively. We estimated the number of genes in the hexaploid wheat genome to range between 94,000 and 96,000 (Supplementary Information, section 2.10). This is reasonably consistent with estimates based on wheat chromosome sequences[13]. Comparing our transcriptome assembly (Supplementary Information, sections 2.8 and 2.9) and wheat harvEST with the wheat OGRs showed that 76% and, respectively, 65% were expressed under the conditions used for RNA isolation. Similar results were found in barley[22], rice[16] and maize[23], indicating that the assemblies are bona fide wheat genes.

We defined the overall extent of gene conservation between wheat and the most closely related sequenced pooid grass, *Brachypodium distachyon*[9,24]. Track 1 of Fig. 2 shows that there is a high degree of overlap between the gene sets of *Brachypodium* and wheat, but with regions of lower conservation, for example on *Brachypodium* chromosomes 1 and 4. Syntenic maps of the *Brachypodium* genome and the A-, B- and D-chromosome groups were created by integrating high-density wheat EST-based markers[25] with *Brachypodium* genes (Fig. 2, tracks 5, 6 and 7, respectively). Supplementary Fig. 7 shows the A-, B- and D-genome markers separately. Syntenic alignments were readily identifiable and conformed to the predicted major patterns[9,26]. We identified many insertions and/or translocations of blocks of genes within the overall conserved patterns of gene order, including the major rearrangement on chromosome 4A as shown on *Brachypodium* chromosome 1 (ref. 20). Lower marker density on the D genome is evident in track 7. The higher-resolution genetic map identified a new syntenic alignment of Triticeae group 5 to *Brachypodium* chromosome 3 genes.

Genome change in polyploid wheat

We determined the influence of polyploidy on gene content in hexaploid wheat by defining the sizes of gene families in hexaploid wheat and the diploid progenitor *Ae. tauschii* from the copy number of genes

Table 2 | Assembly statistics of the orthologous group assembly, the LCG and cDNA assemblies

	Orthologous group assembly* (99% m.i.)	LCG†	cDNA assembly†
Number of sequences	949,279	5,321,847	97,481
Total sequence (bp)	437,512,281	3,800,325,216	93,340,842
Minimum length; maximum length (bp)	79; 7,312	100; 21,721	100; 10,382
N10; N50; N90 (bp)	766; 481; 331	2,234; 884; 420	2,707; 1,325; 509
Mean length (bp)	460.89	714.10	957.53
GC content (%)	48.25	47.69	47.74

* Combined set of 454 sequences that cluster and form contigs and 454 sequences that remain singletons.
† Set of 454 sequences that cluster and form contigs.
bp, base pair.

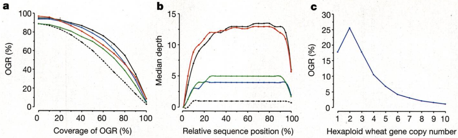


Figure 1 | Coverage of OGRs by wheat 454 sequence reads and simulated 454 reads from rice and maize. **a,** Coverage of OGRs by repeat-masked wheat 454 sequence reads (black line), wheat LCG (black dashed line) and the orthologous group assembly (blue line), together with rice genes (red line) and maize simulated reads (green line). **b,** Median coverage depth over protein-coding regions of OGRs (amino terminus = 0; carboxy terminus = 100). The colour coding is the same as in **a**, except simulated hexaploid reads from rice (red line) were used. **c,** Distribution of wheat gene copy numbers from the orthologous group assembly.

for each OGR, which were then paired with the gene family size of the OGR in sequenced diploid grasses (Supplementary Information, section 2.6). The mean family size was 1.4 members. Supplementary Fig. 8 shows relationships between wheat and diploid orthologous gene family across the full scale of orthologous gene family sizes. This approach accurately reconstructed gene family sizes in simulated maize and 'hexaploid' rice genomes (Figs 3a, b), although larger gene family sizes tended to be underestimated. Figure 3c, d shows the relationships between *Ae. tauschii* and wheat genes. Single-member gene families in hexaploid wheat and *Ae. tauschii* were maintained to a similar extent as those seen in sequenced diploid grasses, consistent with Southern blot analyses of single-copy genes[27]. Using the D genome as a diploid reference, we calculated the Triticeae hexaploid/diploid gene family size ratio to be between 2.5:1 and 2.7:1, derived from the geometric mean (2.5:1) and the slopes of the blue line and the red line (2.7:1) in Fig. 3e. Comparing this with the expected hexaploid/diploid ratio of 3:1 indicates the loss of between 10,000 and 16,000 genes in hexaploid wheat compared with the three diploid progenitors (Supplementary Information, section 2.10). This is consistent with earlier studies of gene loss in newly synthesized wheat polyploids[28] and the erosion of genetic diversity during wheat domestication[6].

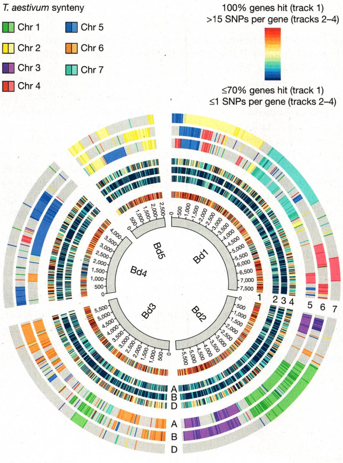


Figure 2 | Alignment of wheat 454 reads, SNPs and genetic maps to the *B. distachyon* genome. The inner circle represent gene order on the five *Brachypodium* chromosomes (Bd1–Bd5). Track 1 illustrates conservation between wheat 454 reads and *Brachypodium* genes, shown as a window of genes present in wheat. Tracks 2–4 show SNP density (the mean number of SNPs per gene in a window of 20 genes) in the A (track 2), B (track 3) and D (track 4) genomes of wheat. Tracks 5–7 show wheat synteny with *Brachypodium* for the A (track 5), B (track 6) and D (track 7) genomes. Genetic markers[25] (shown in darker colours) are colour-coded by wheat chromosome. Gaps between markers are filled in to show synteny (lighter colours).

Despite this overall trend of gene family size reduction, gene families with fewer or more members than expected were identified in *Ae. tauschii* and hexaploid wheat, as shown by green dots (more members) and brown dots (fewer members) in Fig. 3c (*Ae. tauschii*) and Fig. 3d (hexaploid wheat). Supplementary Tables 10–12 show the over- and under-represented functional categories of protein. Most of the over-represented categories in expanded gene families are common to wheat and *Ae. tauschii*: these include ribosome proteins, components of photosystem II, storage proteins, transposon-related proteins, cytochrome P450s, NB-ARC domain proteins involved in defence responses, proteins related to pollen allergens and F-box proteins. Five of the eleven families encoding hydrogen ion transmembrane transporters were significantly more numerous in *Ae. tauschii* than in wheat. Analysis of gene families (Supplementary Fig. 9) showed that they encode different subunits of ATPases. We speculate that they may provide proton gradients to support Na^+ exclusion in *Ae. tauschii*[29] and the accumulation of minerals in other *Aegilops* species[30].

Pseudogene analysis

Several classes of plant DNA transposons[31,32] and retroelements[33] create and amplify gene fragments, disrupt genes and create pseudogenes, which can influence gene expression through epigenetic mechanisms[34]. We identified a set of almost 233,000 gene fragments that mapped to the same regions of their OGRs, forming 'stacks' that were sufficiently divergent not to assemble into their cognate gene assemblies (Fig. 4a). Two classes were identified: those containing Pfam domains and those aligning with non-Pfam domains of OGRs. Nearly 30% of the OGRs had associated gene fragments (Supplementary Table 13) that most frequently covered between 5 and 15% of the OGR length (Fig. 4b). Figure 4c shows that the alignment identities of gene fragments against their OGRs were substantially lower than the identities of cognate regions within wheat gene assemblies. Supplementary Fig. 10 shows the distribution of stacks along genes and the ratio of non-synonymous to synonymous substitutions (K_a/K_s) along the genes. Pfam domains found in stacks were enriched for zinc-finger motifs in mutator transposons (Supplementary Table 14), consistent

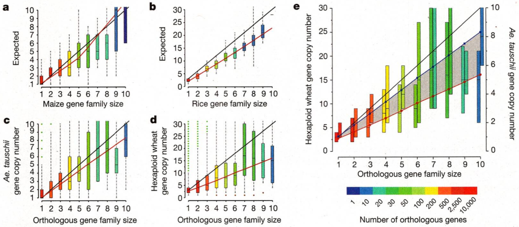


Figure 3 | Gene family sizes in orthologous assemblies of hexaploid wheat, *Ae. tauschii*, simulated maize and hexaploid rice. The boxes and whiskers contain 50% and 90% of the orthologous group assembly genes, respectively. The box colours indicate the number of genes in diploid gene families of different sizes. The black lines represent expected gene family sizes, and the red lines show the gene family sizes determined from the orthologous group assembly, derived by polynomial regression fit. Only gene families with up to ten members are shown. **a,** Maize gene family sizes predicted from orthologous assembly of simulated 454 reads. **b,** Rice gene family sizes predicted from orthologous assembly of simulated 454 reads derived from triplicated rice genes. **c,** *Aegilops tauschii* gene family sizes obtained from orthologous assembly of repeat-masked 454 reads. Expanded gene families are shown as green dots. **d,** Wheat gene family sizes in the orthologous group assembly. **e,** Amalgamation of wheat and *Ae. tauschii* gene copy numbers. The black line shows the respective expected gene copy numbers for wheat and *Ae. tauschii*. The red line shows the regression fit for wheat, and the blue line shows the regression fit for *Ae. tauschii*. The grey zone between these lines estimates the extent of gene loss in hexaploid wheat. For each family size, the left-hand boxes represent hexaploid wheat and the right-hand boxes represent *Ae. tauschii*.

with their role in pseudogene formation[31]. F-box, protein kinase and NB-ARC domains, which are found in the most rapidly evolving gene families in plants[9,35], are also over-represented.

Determining homeologous relationships of gene assemblies

We classified gene assemblies as A-, B- or D-genome-derived according to sequence similarity to Illumina sequence assemblies from *T. monococcum*, cDNA assemblies from *Ae. speltoides* and, respectively, 454 reads from *Ae. tauschii* by applying a support vector machine learning approach (Supplementary Section 5, Supplementary Figs 11 and 12, and Supplementary Tables 15–18). Supplementary Fig. 13 shows that 66% of the gene assemblies were classified with high overall precision (>70%) and recall into the A genome (28.3%), the B genome (29.2%) and the D genome (33.8%). The other 9% of classified assemblies have stop codons. The othes 34% with low classification probabilities are likely to be very similar homeologues. Comparison with a subset of A-, B- and D-genome SNPs confirmed 72% of A-genome classifications and 85% of D-genome classifications (Fig. 2 and Supplementary Table 19). Discrimination of putative B-genome genes was only ~60%, possibly owing both to the use of cDNA sequences for classification when most of the informative sequence polymorphisms are intronic, and to uncertainty about the ancestry of the B genome[5]. The set of 132,552 SNPs allocated to the A, B and D genomes is displayed using *Brachypodium* as a template in tracks 2–4 of Fig. 2.

There were no significant differences between the respective distributions of GO Slim molecular function categories in the A, B and D genes (Supplementary Fig. 14), indicating that at this level of functional categorization there is no biased gene loss[36] in any of the genomes. Nevertheless, analysis of GO Slim terms associated with stop codons in A, B and D gene assemblies showed that there was a strong tendency to retain functional copies of genes encoding transcription factors in all three genomes (Supplementary Fig. 15), similar to the preferential retention of these genes in *Arabidopsis* genome duplications[37]. This indicates that genome-specific transcriptional regulatory networks tend to be maintained in wheat.

Conclusions

Using whole-genome 454 sequencing, we assembled gene sequences representing an essentially complete gene set, and a significant number were assigned to the A, B or D genome. Although the assemblies are fragmentary, they form a powerful framework for identifying genes, accelerating further genome sequencing and facilitating genome-scale analyses. The identification of over 132,000 SNPs in A, B and D genes facilitates analysis of quantitative trait loci and association studies of

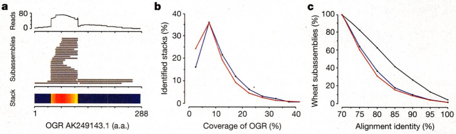


Figure 4 | Pseudogene identification and analysis. a, Visualization of an OGR and associated wheat sequences. The top track shows the hit count profile of mapped 454 reads. The lower tracks show subassemblies of three wheat genes and a stacked region of gene fragments. Read depth is represented by the heat map. **b,** Coverage of the OGR by Pfam-containing gene fragments and pseudogenes. The blue and red lines represent stacks with and without protein domains, respectively. **c.** Protein identity between subassemblies forming stacks of gene fragments. The blue and red lines represent stacks with and without protein domains, respectively, and the black line represents subassemblies forming genes.

traits. Comparison with the sequences of diploid progenitors and relatives showed pronounced reductions in the size of large gene families in wheat despite the relatively recent formation of the hexaploid (Fig. 3e), consistent with smaller-scale analyses[28,38]. The scale of gene loss in hexaploid wheat compared with maize[36] and *Brassica rapa*[39] is significantly smaller, possibly as a result of its relatively recent origin and the absence of intergenome recombination[40]. Nevertheless, gene loss in wheat could be rapid, as shown in the newly created allopolyploid *Tragopogon miscellus*[41]. Most functional classes show equal gene loss in the three genomes, but families of transcription factors showed a clear tendency to be retained as functional genes in all three genomes. These may maintain transcriptional networks in each genome and contribute to non-additive gene expression[42] and genome plasticity. In contrast to the overall loss of gene family members, several classes of gene families with predicted roles in defence, nutritional content, energy metabolism and growth have increased sizes in the Triticeae lineage, possibly as a result of selection during domestication.

Major efforts are underway to improve wheat productivity by increasing genetic diversity in breeding materials and through genetic analysis of traits[43]. The genomic resources that we have developed promise to accelerate progress by facilitating the identification of useful variation in genes of wheat landraces and progenitor species, and by providing genomic landmarks to guide progeny selection. Analysis of complex polygenic traits such as yield and nutrient use efficiency will also be accelerated, contributing to sustainable increases in wheat crop production.

METHODS SUMMARY

A single-seed descent line of *T. aestivum* landrace Chinese Spring was sequenced, because it is widely used for cytogenetic analysis[44] and physical mapping[15]. *Triticum monococcum* accession 4342-96 is a community standard line for targeting induced local lesions in genomes, physical mapping and genetic analysis; and *Ae. tauschii* ssp *strangulata* accession AL8/78, which is used for physical and genetic mapping, was sequenced using 454 technology.

Sequence for the *T. aestivum* wheat gene assembly was generated using Roche 454 pyrosequencing on the GS FLX Titanium and GS FLX+ platforms. Additional sequence read data sets for *T. aestivum*, *T. monococcum* and *Ae. tauschii* were generated using three platforms, Illumina, 454 and SOLiD, to analyse homeologous sequences and SNPs (a list of all data sets is in Supplementary Table 2). Orthologous groups were created from rice, sorghum and *B. distachyon* genome sequences and barley full-length cDNA sequences. Wheat gene assemblies were named according to their OGR and were identified by a seven-digit identifier and their predicted genome (for example Traes_Bradi1g12345_0000001_D and Traes_Sb3g33333_6543210_A). Gene and cDNA assemblies can be searched at the MIPS Wheat Genome Database (http://mips.helmholtz-muenchen.de/plant/wheat/uk454survey/index.jsp). All sequence data has been deposited in publicly accessible databases, described in Supplementary Information. Sequence assemblies, annotated gene sequences and their relationships are available for download from the European Bioinformatics Institute (www.ebi.ac.uk) and viewing in a synteny-based Ensembl genome browser. Annotated gene sequences and their relationships can be viewed in a *Brachypodium* synteny-based Ensembl genome browser (http://plants.ensembl.org/brachypodium_distachyon).

Received 4 March; accepted 1 October 2012.

1. United States Department of Agriculture. *World Agricultural Supply and Demand Estimates*. Report No. WASDE-511; http://usda01.library.cornell.edu/usda/current/wasde/wasde-10-11-2012.pdf (2012).
2. Food and Agriculture Organisation of the United Nations. http://faostat.fao.org/default.aspx?lang=en (2011).
3. Nesbitt, M. & Samuel, D. in *Hulled Wheats* (eds Padulosi, S., Hammer, K. & Heller, J.) 41–100 (Proc. 1st Internat. Workshop Hulled Wheats, International Plant Genetic Resources Institute, 1996).
4. Dvorak, J., Akhunov, E. D., Akhunov, A. R., Deal, K. R. & Luo, M. C. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Mol. Biol. Evol.* **23,** 1386–1396 (2006).
5. Salse, J. *et al.* New insights into the origin of the B genome of hexaploid wheat: evolutionary relationships at the SPA genomic region with the S genome of the diploid relative Aegilops speltoides. *BMC Genomics* **9,** 555 (2008).
6. Haudry, A. *et al.* Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* **24,** 1506–1517 (2007).
7. Moore, G., Devos, K. M., Wang, Z. & Gale, M. D. Cereal genome evolution. Grasses, line up and form a circle. *Curr. Biol.* **5,** 737–739 (1995).
8. Paterson, A. H. *et al.* The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457,** 551–556 (2009).
9. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463,** 763–768 (2010).
10. Smith, D. B. & Flavell, R. B. Characterisation of the wheat genome by association genetics. *Chromosoma* **50,** 223–242 (1975).
11. Baucom, R. S. *et al.* Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet.* **5,** e1000732 (2009).
12. Wicker, T. *et al.* Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell* **23,** 1706–1718 (2011).
13. Choulet, F. *et al.* Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* **22,** 1686–1701 (2010).
14. Gill, B. S. *et al.* A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. *Genetics* **168,** 1087–1096 (2004).
15. Paux, E. *et al.* A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* **322,** 101–104 (2008).
16. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436,** 793–800 (2005).
17. Li, L., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13,** 2178–2189 (2003).
18. Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. & Shinozaki, K. TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol.* **150,** 1135–1146 (2009).
19. Richter, D. C., Ott, F., Auch, A. F., Schmid, R. & Huson, D. H. MetaSim: a sequencing simulator for genomics and metagenomics. *PLoS ONE* **3,** e3373 (2008).
20. Hernandez, P. *et al.* Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.* **69,** 377–386 (2012).
21. Massa, A. N. *et al.* Gene space dynamics during the evolution of Aegilops tauschii, Brachypodium distachyon, Oryza sativa, and Sorghum bicolor genomes. *Mol. Biol. Evol.* **28,** 2537–2547 (2011).
22. The International Barley Genome Sequencing Consortium. A physical, genetic, and functional sequence assembly of the barley genome. *Nature* doi:10.1038/nature11543 (this issue).
23. Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326,** 1112–1115 (2009).
24. Lee, E. K. *et al.* A functional phylogenomic view of the seed plants. *PLoS Genet.* **7,** e1002411 (2011).
25. Allen, A. M. *et al.* Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). *Plant Biotechnol. J.* **9,** 1086–1099 (2011).
26. Salse, J. *et al.* Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* **20,** 11–24 (2008).
27. Qi, L. L. *et al.* A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* **168,** 701–712 (2004).
28. Ozkan, H., Levy, A. A. & Feldman, M. Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops-Triticum) group. *Plant Cell* **13,** 1735–1747 (2001).
29. Shavrukov, Y., Langridge, P. & Tester, M. Salinity tolerance and sodium exclusion in genus Triticum. *Breed. Sci.* **59,** 671–678 (2009).
30. Wang, S., Yin, L., Tanaka, K., Tanaka, H. & Tsujimoto, H. Wheat-Aegilops chromosome addition lines showing high iron and zinc contents in grains. *Breed. Sci.* **61,** 189–195 (2011).
31. Jiang, N., Bao, Z., Zhang, X., Eddy, S. R. & Wessler, S. R. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431,** 569–573 (2004).
32. Morgante, M. *et al.* Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genet.* **37,** 997–1002 (2005).
33. Jin, Y. K. & Bennetzen, J. L. Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bs1 retroelement of maize. *Plant Cell* **6,** 1177–1186 (1994).
34. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430,** 471–476 (2004).
35. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815 (2000).
36. Schnable, J. C., Springer, N. M. & Freeling, M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl Acad. Sci. USA* **108,** 4069–4074 (2011).
37. Maere, S. *et al.* Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA* **102,** 5454–5459 (2005).
38. Gu, Y. Q., Coleman-Derr, D., Kong, X. & Anderson, O. D. Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four Triticeae genomes. *Plant Physiol.* **135,** 459–470 (2004).
39. Mun, J. H. *et al.* Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* **10,** R111 (2009).
40. Riley, R. Genetic control of cytologically diploid behaviour of hexaploid wheat. *Nature* **182,** 713–715 (1958).
41. Buggs, R. J. *et al.* Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol.* **22,** 248–252 (2012).

42. Pumphrey, M., Bai, J., Laudencia-Chingcuanco, D., Anderson, O. & Gill, B. S. Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics* **181,** 1147–1157 (2009).
43. Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **327,** 818–822 (2010).
44. Sears, E. R. in *Chromosome Manipulation and Plant Genetics* (eds Riley, R. & Lewis, K. R.) 22–45 (Oliver and Boyd, 1966).

Supplementary Information is available in the online version of the paper.

Acknowledgements DNA sequence was generated by The University of Liverpool Centre for Genomic Research (United Kingdom), 454 Life Sciences (United States), The Cold Spring Harbor Woodbury Genome Centre (United States) and The Genome Analysis Centre (United Kingdom). This work was supported by UK Biological and Biotechnological Sciences Research Council (BBSRC) grants BB/G012865, BB/G013985/1 and BB/G013004/1, to K.J.E., M.W.B. and N. Hall; a Wolfson Merit Award from the Royal Society, to N. Hall; BBSRC Strategic Programme grant B/J004588/1 (GRO), to M.W.B.; EC TriticeaeGenome grant number 212019, to K.F.X.M. and M.W.B.; The TRITEX Project of the Plant20130 Initiative of the German Ministry of Education and Research grant number 0315954C, to K.F.X.M.; EC Transplant Grant 283496, to K.F.X.M. and P.K., a BBSRC Career Development Fellowship BB/H022333/1, to A.H., US NSF grants IOS-1032105 and DBI-0923128, to W.R.M.; USDA-NIFA grant 2008-35300-04588, to B.G.; and US NSF grants DBI-0701916, to J.D., and DBI-0822100, to S. Kianian.

Author Contributions R.B., M.S., M.P., G.L.A.B. and R.D. are joint first authors. K.J.E., M.W.B., N. Hall and A.H. designed the project; W.R.M., M.K., M.T., I.B., J.D., M.-C.L., O.A., S. Kianian, N. Huo, B.G. and S.S. provided data and advice; R.D., N.M. and S. Kay conducted experiments; K.F.X.M., N. Hall and M.W.B. planned and conducted analyses; and R.B., M.S., M.P., G.L.A.B., A.M.A., D.B., D.W., P.K. and A.H. carried out analyses. K.J.E., A.H., W.R.M. and R.B. contributed to the text and M.W.B., N. Hall and K.F.X.M. wrote the manuscript. All authors commented on the manuscript.

Author Information Sequence assemblies have been submitted to the European Nucleotide Archive under project accession number PRJEB568. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.B. (michael.bevan@jic.ac.uk), K.F.X.M. (k.mayer@helmholtz-muenchen.de), N. Hall (Neil.Hall@liverpool.ac.uk), A. H. (Anthony.Hall@liverpool.ac.uk), or K.J.E. (k.j.edwards@bristol.ac.uk). This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-sa/3.0/

doi:10.1038/nature11543

A physical, genetic and functional sequence assembly of the barley genome

The International Barley Genome Sequencing Consortium*

Barley (*Hordeum vulgare* L.) is among the world's earliest domesticated and most important crop plants. It is diploid with a large haploid genome of 5.1 gigabases (Gb). Here we present an integrated and ordered physical, genetic and functional sequence resource that describes the barley gene-space in a structured whole-genome context. We developed a physical map of 4.98 Gb, with more than 3.90 Gb anchored to a high-resolution genetic map. Projecting a deep whole-genome shotgun assembly, complementary DNA and deep RNA sequence data onto this framework supports 79,379 transcript clusters, including 26,159 'high-confidence' genes with homology support from other plant genomes. Abundant alternative splicing, premature termination codons and novel transcriptionally active regions suggest that post-transcriptional processing forms an important regulatory layer. Survey sequences from diverse accessions reveal a landscape of extensive single-nucleotide variation. Our data provide a platform for both genome-assisted research and enabling contemporary crop improvement.

Cultivated barley, derived from its wild progenitor *Hordeum vulgare* ssp. *spontaneum*, is among the world's earliest domesticated crop species[1] and today represents the fourth most abundant cereal in both area and tonnage harvested (http://faostat.fao.org). Approximately three-quarters of global production is used for animal feed, 20% is malted for use in alcoholic and non-alcoholic beverages, and 5% as an ingredient in a range of food products[2]. Barley is widely adapted to diverse environmental conditions and is more stress tolerant than its close relative wheat[3]. As a result, barley remains a major food source in poorer countries[4], maintaining harvestable yields in harsh and marginal environments. In more developed societies it has recently been classified as a true functional food. Barley grain is particularly high in soluble dietary fibre, which significantly reduces the risk of serious human diseases including type II diabetes, cardiovascular disease and colorectal cancers that afflict hundreds of millions of people worldwide[5]. The USA Food and Drug Administration permit a human health claim for cell-wall polysaccharides from barley grain.

As a diploid, inbreeding, temperate crop, barley has traditionally been considered a model for plant genetic research. Large collections of germplasm containing geographically diverse elite varieties, landraces and wild accessions are readily available[6] and undoubtedly contain alleles that could ameliorate the effect of climate change and further enhance dietary fibre in the grain. Enriching its broad natural diversity, extensive characterized mutant collections containing all of the morphological and developmental variation observed in the species have been generated, characterized and meticulously maintained. The major impediment to the exploitation of these resources in fundamental and breeding science has been the absence of a reference genome sequence, or an appropriate enabling alternative. Providing either of these has been the primary research challenge to the global barley community.

In response to this challenge, we present a novel model for delivering the genome resources needed to reinforce the position of barley as a model for the Triticeae, the tribe that includes bread and durum wheats, barley and rye. We introduce the barley genome gene space, which we define as an integrated, multi-layered informational resource that provides access to the majority of barley genes in a highly structured physical and genetic framework. In association with comparative sequence and transcriptome data, the gene space provides a new molecular and cellular insight into the biology of the species, providing a platform to advance gene discovery and genome-assisted crop improvement.

A sequence-enriched barley physical map

We constructed a genome-wide physical map of the barley cultivar (cv.) Morex by high-information-content fingerprinting[7] and contig assembly[8] of 571,000 bacterial artificial chromosome (BAC) clones (~14-fold haploid genome coverage) originating from six independent BAC libraries[9]. After automated assembly and manual curation, the physical map comprised 9,265 BAC contigs with an estimated N50 contig size of 904 kilobases and a cumulative length of 4.98 Gb (Methods, Supplementary Note 2). It is represented by a minimum tiling path (MTP) of 67,000 BAC clones. Given a genome size of 5.1 Gb[10], more than 95% of the barley genome is represented in the physical map, comparing favourably to the 1,036 contigs that represent 80% of the 1 Gb wheat chromosome 3B[11].

We enhanced the physical map by integrating shotgun sequence information from 5,341 gene-containing[12,13] and 937 randomly selected BAC clones (Methods, Supplementary Notes 2 and 3, and Supplementary Table 4), and 304,523 BAC-end sequence (BES) pairs (Supplementary Table 3). These provided 1,136 megabases (Mb) of genomic sequence integrated directly into the physical map (Supplementary Tables 3 and 4). This framework facilitated the incorporation of whole-genome shotgun sequence data and integration of the physical and genetic maps. We generated whole-genome shotgun sequence data from genomic DNA of cv. 'Morex' by short-read Illumina GAIIx technology, using a combination of 300 base pairs (bp) paired-end and 2.5 kb mate-pair libraries, to >50-fold haploid genome coverage (Supplementary Note 3.3). *De novo* assembly resulted in sequence contigs totalling 1.9 Gb. Due to the high proportion of repetitive DNA, a substantial part of the whole-genome shotgun data collapsed into relatively small contigs characterized by exceptionally high read depths. Overall, 376,261 contigs were larger than 1 kb (N50 = 264,958 contigs, N50 length = 1,425 bp). Of these, 112,989

---

*A list of authors and their affiliations appears at the end of the paper.

(308 Mb) could be anchored directly to the sequence-enriched physical map by sequence homology.

We implemented a hierarchical approach to further anchor the physical and genetic maps (Methods, Supplementary Note 4). A total of 3,241 genetically mapped gene-based single-nucleotide variants (SNV) and 498,165 sequence-tag genetic markers[14] allowed us to use sequence homology to assign 4,556 sequence-enriched physical map contigs spanning 3.9 Gb to genetic positions along each barley chromosome. An additional 1,881 contigs were assigned to chromosomal bins by sequence homology to chromosome-arm-specific sequence data sets[15] (Supplementary Note 4.4). Thus, 6,437 physical map contigs totalling 4.56 Gb (90% of the genome), were assigned to chromosome arm bins, the majority in linear order. Non-anchored contigs were typically short and lacked genetically informative sequences required for positional assignment.

Consistent with genome sequences of other grass species[16] the pericentromeric and centromeric regions of barley chromosomes exhibit significantly reduced recombination frequency, a feature that compromises exploitation of genetic diversity and negatively impacts genetic studies and plant breeding. Approximately 1.9 Gb or 48% of the genetically anchored physical map (3.9 Gb) was assigned to these regions (Fig. 1 and Supplementary Fig. 11).

Repetitive nature of the barley genome

A characteristic of the barley genome is the abundance of repetitive DNA[17]. We observed that approximately 84% of the genome is comprised of mobile elements or other repeat structures (Supplementary Note 5). The majority (76% in random BACs) of these consists of retrotransposons, 99.6% of which are long terminal repeat (LTR) retrotransposons. The non-LTR retrotransposons contribute only 0.31% and the DNA transposons 6.3% of the random BAC sequence. In the fraction of the genome with a high proportion of repetitive elements, the LTR *Gypsy* retrotransposon superfamily was 1.5-fold more abundant than the *Copia* superfamily, in contrast to observations in both *Brachypodium*[18] and rice[19]. However, gene-bearing BACs were slightly depleted of retrotransposons, consistent with *Brachypodium*[18] where young *Copia* retroelements are preferentially found in gene-rich, recombinogenic regions from which inactive *Gypsy* retroelements have been lost by LTR–LTR recombination. Overall, we see reduced repetitive DNA content within the terminal 10% of the physical map of each barley chromosome arm (Fig. 1). Class I and II elements show non-quantitative reverse-image distribution along barley chromosomes (Fig. 1), a feature shared with other grass genomes[16,20] and shown by fluorescence *in situ* hybridization (FISH) mapping[17]. Not surprisingly, the whole-genome shotgun assembly shows a lower abundance of LTR retrotransposons (average 53%) than gene-bearing BACs. That LTR retrotransposons are long (~10 kb), highly repetitive and often nested[21] supports our assumption that short reads either collapsed or did not assemble. Short interspersed elements (SINEs)[22], short (80–600 bp) non-autonomous retrotransposons that are highly repeated in barley, showed no differential exclusion from the assemblies. However, miniature inverted-repeat transposable elements (MITEs), small non-autonomous DNA transposons[23], were twofold enriched in the whole-genome shotgun assemblies compared with BES reads or random BACs, consistent with the gene richness of the assemblies and their association with genes[23]. Both MITEs and SINEs are 1.5 to 2-fold enriched in gene-bearing BACs which could indicate that SINEs are also preferentially integrated into gene-rich regions, or because they are older than LTR retroelements, may simply remain visible in and around genes where retro insertions have been selected against.

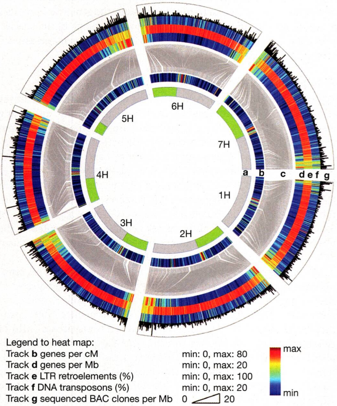


Figure 1 | Landscape of the barley gene space. Track **a** gives the seven barley chromosomes. Green/grey colour depicts the agreement of anchored fingerprint (FPC) contigs with their chromosome arm assignment based on chromosome-arm-specific shotgun sequence reads (for further details see Supplementary Note 4). For 1H only whole-chromosome sequence assignment was available. Track **b**, distribution of high-confidence genes along the genetic map; track **c**, connectors relate gene positions between genetic and the integrated physical map given in track **d**. Position and distribution of track **e** class I LTR-retroelements and track **f** class II DNA transposons are given. Track **g**, distribution and positioning of sequenced BACs.

Transcribed portion of the barley genome

The transcribed complement of the barley gene space was annotated by mapping 1.67 billion RNA-seq reads (167 Gb) obtained from eight stages of barley development as well as 28,592 barley full-length cDNAs[24] to the whole-genome shotgun assembly (Methods, Supplementary Notes 6, 7 and Supplementary Tables 20–22). Exon detection and consensus gene modelling revealed 79,379 transcript clusters, of which 75,258 (95%) were anchored to the whole-genome shotgun assembly (Supplementary Notes 7.1.1 and 7.1.2). Based on a gene-family-directed comparison with the genomes of *Sorghum*, rice, *Brachypodium* and *Arabidopsis*, 26,159 of these transcribed loci fall into clusters and have homology support to at least one reference genome (Supplementary Fig. 16); they were defined as high-confidence genes. Comparison against a data set of metabolic genes in *Arabidopsis thaliana*[25] indicated a detection rate of 86%, allowing the barley gene-set to be estimated as approximately 30,400 genes. Due to lack of homology and missing support from gene family clustering, 53,220 transcript loci were considered low-confidence (Table 1). High-confidence and low-confidence barley genes exhibited distinct characteristics: 75% of the high-confidence genes had a multi-exon structure, compared with only 27% of low-confidence genes (Table 1). The mean size of high-confidence genes was 3,013 bp compared with 972 bp for low-confidence genes. A total of 14,481 low-confidence genes showed distant homology to plant proteins in public databases (Supplementary Notes 7.1.2, 7.1.4 and Supplementary Fig. 18), identifying them as potential gene fragments known to populate Triticeae genomes at high copy number and that often result from transposable element activity[26].

A total of 15,719 high-confidence genes could be directly associated with the genetically anchored physical map (Supplementary Note 4). An additional 3,743 were integrated by invoking a conservation of synteny model (Supplementary Note 4.5) and a further 4,692 by association

Table 1 | Characteristics of high-confidence and low-confidence gene sets in barley

	High confidence	Low confidence
Number of genes	26,159	53,220
Gene loci positioned on barley cultivar Morex assembly*	24,243 (93%)	51,015 (96%)
Single exon	5,954 (25%)	37,395 (73%)
Multi exon	18,289 (75%)	13,620 (27%)
Number of distinct exons†	184,710	107,768
Mean number of distinct exons per gene	7.62	2.11
Number of genes with alternative transcript variants	13,299 (55%)	8,214 (16%)
Total number of predicted transcripts	62,426	69,266
Mean number of transcripts per gene	2.58	1.36
Mean gene locus size (first to last exon)	3,013 bp	972 bp
Mean transcript size (UTR, CDS)	1,878bp	931 bp
Mean exon size	454 bp	536 bp
Gene loci not positioned on barley cv. Morex assembly‡	1,916 (7%)	2,205 (4%)
Tagged by unmapped RNA-seq reads	1,657 (86%)	1,127 (51%)
Not tagged by unmapped RNA-seq reads	259 (14%)	1,078 (49%)

* Gene locus representatives are (1) RNA-seq based transcript or (2) barley fl-cDNA that were mapped to the barley cultivar Morex assembly or tagged by RNA-seq based transcript during clustering.
† Exons of two or more transcripts were counted once if they have identical start and stop positions.
‡ Gene locus representatives are barley fl-cDNAs that were not mapped to the barley cultivar Morex assembly and not matched by any RNA-seq based transcript
CDS, coding sequence.

with chromosome arm whole-genome shotgun data (Supplementary Note 4.4 and Supplementary Table 15). Importantly, the N50 length of whole-genome shotgun sequence contigs containing high-confidence genes was 8,172 bp, which is generally sufficient to include the entire coding sequence, and 5′ and 3′ untranslated regions (UTRs). Overall 24,154 high-confidence genes (92.3%) were associated and positioned in the physical/genetic scaffold, representing a gene density of five genes per Mb. Proximal and distal ends of chromosomes are more gene-rich, on average containing 13 genes per Mb (Fig. 1).

In comparison with sequenced model plant genomes, gene family analysis (Supplementary Note 7.1.3) revealed some gene families that exhibited barley-specific expansion. We defined the functions of members of these families using gene ontology (GO) and PFAM protein motifs (Supplementary Table 25). Gene families with highly overrepresented GO/PFAM terms included genes encoding (1,3)-β-glucan synthases, protease inhibitors, sugar-binding proteins and sugar transporters. NB-ARC (a nucleotide-binding adaptor shared by APAF-1, certain R gene products and CED-4[27]) domain proteins, known to be involved in defence responses, were also overrepresented, including 191 NBS-LRR type genes. These tended to cluster towards the distal regions of barley chromosomes (Supplementary Fig. 17), including a major group on barley chromosome 1HS, co-localizing with the *MLA* powdery mildew resistance gene cluster[28]. Biased allocation to recombination-rich regions provides the genomic environment for generating sequence diversity required to cope with dynamic pathogen populations[29,30]. It is noteworthy that the highly over-represented (1,3)-β-glucan synthase genes have also been implicated in plant–pathogen interactions[31].

Regulation of gene expression

Deep RNA sequence data (RNA-seq) provided insights into the spatial and temporal regulation of gene expression (Supplementary Note 7.2). We found 72–84% of high-confidence genes to be expressed in all spatiotemporal RNA-seq samples (Fig. 2a), slightly lower than reported for rice[32] where ~95% of transcripts were found in more than one developmental or tissue sample. More importantly, 36–55% of high-confidence barley genes seemed to be differentially regulated between samples (Fig. 2b), highlighting the inherent dynamics of barley gene expression.

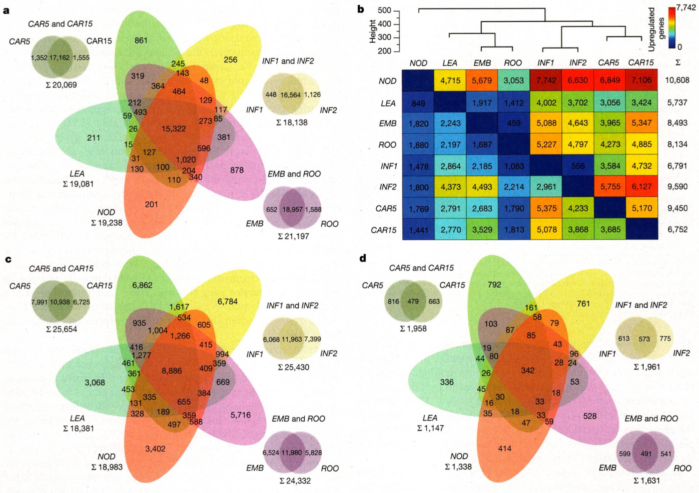
Two notable features support the importance of post-transcriptional processing as a central regulatory layer (Supplementary Notes 7.3 and 7.4). First, we observed evidence for extensive alternative splicing. Of the intron-containing high-confidence barley genes, 73% had evidence of alternative splicing (55% of the entire high-confidence set). The spatial and temporal distribution of alternative splicing transcripts deviated significantly from the general occurrence of transcripts in the different tissues analysed (Fig. 2c). Only 17% of alternative splicing transcripts were shared among all samples, and 17–27% of the alternative splicing transcripts were detected only in individual samples, indicating pronounced alternative splicing regulation. We found 2,466 premature termination codon-containing (PTC+) alternative splicing transcripts (9.4% of high-confidence genes) (Fig. 2d and Table 2), similar to the percentage of nonsense-mediated decay (NMD)-controlled genes in a wide range of species[33,34]. Premature termination codons activate the NMD pathway[35], which leads to rapid degradation of PTC+ transcripts, and have been associated with transcriptional regulation during disease and stress response in human and *Arabidopsis*, respectively[34,36–39]. The distribution of PTC+ transcripts was strikingly dissimilar, both spatially and temporally, with only 7.4% shared and between 31% and 40% exclusively observed in only a single sample (Fig. 2d). Genes encoding PTC+-containing transcripts show a broad spectrum of GO terms and PFAM domains and are more prevalent in expanded gene families. These observations support a central role for alternative splicing/NMD-dependent decay of PTC+ transcripts as a mechanism that controls the expression of many different barley genes.

Second, recent reports have highlighted the abundance of novel transcriptionally active regions in rice that lack homology to protein-coding genes or open reading frames (ORFs)[40]. In barley as many as 27,009 preferentially single-exon low-confidence genes can be classified as putative novel transcriptionally active regions (Supplementary Note 7.1.4). We investigated their potential significance by comparing the homology of barley novel transcriptionally active regions with the rice and *Brachypodium* genomes that respectively represent 50 and 30 million years of evolutionary divergence[18]. A total of 4,830 and 2,450 novel transcriptionally active regions yielded a homology match to the *Brachypodium* and rice genomes, respectively (intersection of 2,046; BLAST P value $\leq 10^{-5}$), indicating a putative functional role in pre-mRNA processing or other RNA regulatory processes[41,42].

Natural diversity

Barley was domesticated approximately 10,000 years ago[1]. Extensive genotypic analysis of diverse germplasm has revealed that restricted outcrossing (0–1.8%)[43], combined with low recombination in pericentromeric regions, has resulted in modern germplasm that shows limited regional haplotype diversity[44]. We investigated the frequency and distribution of genome diversity by survey sequencing four diverse barley cultivars ('Bowman', 'Barke', 'Igri' and 'Haruna Nijo') and an *H. spontaneum* accession (Methods and Supplementary Note 8) to a depth of 5–25-fold coverage, and mapping sequence reads against the barley cultivar 'Morex' gene space. We identified more than 15 million non-redundant single-nucleotide variants (SNVs). *H. spontaneum* contributed almost twofold more SNV than each of the cultivars (Supplementary Table 28). Up to 6 million SNV per accession could be assigned to chromosome arms, including up to 350,000 associated with exons (Supplementary Table 29). Approximately 50% of the exon-located SNV were integrated into the genetic/physical framework (Fig. 3, Supplementary Table 30 and Supplementary Fig. 31), providing a platform to establish true genome-wide marker technology for high-resolution genetics and genome-assisted breeding.

We observed a decrease in SNV frequency towards the centromeric and peri-centromeric regions of all barley chromosomes, a pattern that seemed more pronounced in the barley cultivars. This trend was supported by SNV identified in RNA-seq data from six additional cultivars mapped onto the Morex genomic assembly (Supplementary Note 8.2). We attribute this pattern of eroded genetic diversity to low recombination in the pericentromeric regions, which reduces effective population size and consequently haplotype diversity. Whereas



	NOD	LEA	EMB	ROO	INF1	INF2	CAR5	CAR15	Σ
NOD		4,715	5,679	3,053	7,742	6,630	6,849	7,106	10,608
LEA	849		1,917	1,412	4,002	3,702	3,056	3,424	5,737
EMB	1,820	2,243		459	5,088	4,643	3,965	5,347	8,493
ROO	1,880	2,197	1,687		5,227	4,797	4,273	4,885	8,134
INF1	1,478	2,864	2,185	1,083		556	3,584	4,732	6,791
INF2	1,800	4,373	4,493	2,214	2,961		5,755	6,127	9,590
CAR5	1,769	2,791	2,683	1,790	5,375	4,233		5,170	9,450
CAR15	1,441	2,770	3,529	1,813	5,078	3,868	3,685		6,752

Figure 2 | Atlas of barley gene expression. **a**, Barley gene expression in different spatial and temporal RNA-seq samples (Supplementary Notes 6, 7). Numbers refer to high-confidence genes. **b**, Dendrogram depicting relatedness of samples and colour-coded matrix showing number of significantly upregulated high-confidence genes in pairwise comparisons. Σ, total number of non-redundant high-confidence genes upregulated in comparison to all other samples. Height, complete linkage cluster distance ($\log_2$(fragments per kilobase of exon per million fragments mapped)); see Supplementary Note 7.2.5.1. **c**, Distribution and overlap of alternatively spliced barley transcripts between RNA-seq samples. **d**, Distribution and overlap of alternative splicing transcripts fulfilling criteria for PTC+ as detected in different spatial and temporal RNA-seq samples (Supplementary Note 7.4).

Table 2 | Alternative splicing and transcripts containing PTCs in high-confidence genes

General statistics of alternative splicing in high-confidence genes	
High-confidence genes with RNA-seq data to monitor alternative splicing	24,243
Predicted transcripts at high-confidence genes	62,426
Transcripts with complete CDS structures*	62,256
Transcripts with partial CDS structures†	170
Genes with alternative transcripts	13,299
Predicted transcripts derived from genes with alternative splicing	51,482
Premature stop codon analysis	
Predicted transcripts used for PTC analysis‡	51,338
Transcripts without PTC	41,461 (81%)
Transcripts containing PTC	9,877
PTC caused by intron retention	5,286 (10%)
PTC+ transcripts predicted to be NMD****-sensitive	4,591 (9%)
Gene loci incorporating PTC+/NMD transcripts	2,466

* Entire predicted coding sequence (100%) was transferred to transcript model on barley cultivar Morex contigs.
† Predicted coding sequence could not be completely projected to genomic transcript model (partial mapping of fl-cDNA).
‡ Only transcripts with structures for entire coding sequence on barley cultivar Morex WGS assembly were considered.
CDS, coding sequence.

H. spontaneum may serve here as a reservoir of genetic diversity, using this diversity may itself be compromised by restricted recombination and the consequent inability to disrupt tight linkages between desirable and deleterious alleles. Surprisingly, the short arm of chromosome 4H had a significantly lower SNV frequency than all other barley chromosomes (Supplementary Fig. 33). This may be a consequence of a further reduction in recombination frequency on this chromosome, which is genetically (but not physically) shortest. Reduced SNV diversity was also observed in regions we interpret to be either the consequences of recent breeding history or could indicate landmarks of domestication (Fig. 3).

Discussion

The size of Triticeae cereal genomes, due to their highly repetitive DNA composition, has severely compromised the assembly of whole-genome shotgun sequences and formed a barrier to the generation of high-quality reference genomes. We circumvented these problems by integrating complementary and heterogeneous sequence-based genomic and genetic data sets. This involved coupling a deep physical map with high density genetic maps, superimposing deep short-read whole-genome shotgun assemblies, and annotating the resulting linear, albeit punctuated, genomic sequence with deep-coverage RNA-derived data (full-length cDNA and RNA-seq). This allowed us to systematically delineate approximately 4 Gb (80%) of the genome, including more than 90% of the expressed genes. The resulting genomic framework

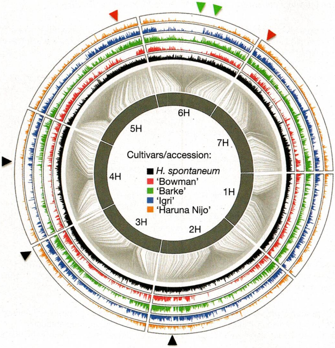


Figure 3 | Single nucleotide variation (SNV) frequency in barley. Barley chromosomes indicated as inner circle of grey bars. Connector lines give the genetic/physical relationship in the barley genome. SNV frequency distribution displayed as five coloured circular histograms (scale, relative abundance of SNVs within accession; abundance, total number of SNVs in non-overlapping 50-kb intervals of concatenated 'Morex' genomic scaffold; range, zero to maximum number of SNVs per 50-kb interval). Selected patterns of SNV frequency indicated by coloured arrowheads (for further details see Supplementary Note 8). Colouring of arrowheads refers to cultivar with deviating SNV frequency for the respective region.

provides a detailed insight into the physical distribution of genes and repetitive DNA and how these features relate to genetic characteristics such as recombination frequency, gene expression and patterns of genetic variation.

The centromeric and peri-centromeric regions of barley chromosomes contain a large number of functional genes that are locked into recombinationally 'inert' genomic regions[45,46]. The gene-space distribution highlights that these regions expand to almost 50% of the physical length of individual chromosomes. Given well-established levels of conserved synteny, this will probably be a general feature of related grass genomes that will have important practical implications. For example, infrequent recombination could function to maintain evolutionarily selected and co-adapted gene complexes. It will certainly restrict the release of the genetic diversity required to decouple advantageous from deleterious alleles, a potential key to improving genetic gain. Understanding these effects will have important consequences for crop improvement. Moreover, for gene discovery, forward genetic strategies based on recombination will not be effective in these regions. Whereas alternative approaches exist for some targets (for example, by coupling resequencing technologies with collections of natural or induced mutant alleles), for most traits it remains a serious impediment. Some promise may lie in manipulating patterns of recombination by either genetic or environmental intervention[47]. Quite strikingly, our data also reveal that a complex layer of post-transcriptional regulation will need to be considered when attempting to link barley genes to functions. Connections between post-transcriptional regulation such as alternative splicing and functional biological consequences remain limited to a few specific examples[48], but the scale of our observations suggest this list will expand considerably.

In conclusion, the barley gene space reported here provides an essential reference for genetic research and breeding. It represents a hub for trait isolation, understanding and exploiting natural genetic diversity and investigating the unique biology and evolution of one of the world's first domesticated crops.

METHODS SUMMARY

Methods are available in the online version of the paper.

Full Methods and any associated references are available in the online version of the paper.

Received 2 May; accepted 30 August 2012.

Published online 17 October 2012.

1. Purugganan, M. D. & Fuller, D. Q. The nature of selection during plant domestication. *Nature* **457,** 843–848 (2009).
2. Blake, T., Blake, V., Bowman, J. & Abdel-Haleem, H. in *Barley: Production, Improvement and Uses* (ed. S. E. Ullrich) 522–531 (Wiley-Blackwell, 2011).
3. Nevo, E. *et al.* Evolution of wild cereals during 28 years of global warming in Israel. *Proc. Natl Acad. Sci. USA* **109,** 3412–3415 (2012).
4. Grando, S. & Macpherson, H. G. in *Proceedings of the International Workshop on Food Barley Improvement,* 14–17 January 2002, Hammamet, Tunisia 156 (ICARDA, Aleppo, Syria, 2005).
5. Collins, H. M. *et al.* Variability in fine structures of noncellulosic cell wall polysaccharides from cereal grains: potential importance in human health and nutrition. *Cereal Chem.* **87,** 272–282 (2010).
6. Bockelman, H. E. & Valkoun, J. in *Barley: Production, Improvement, and Uses* (ed. S. E. Ullrich) 144–159 (Wiley-Blackwell, 2011).
7. Luo, M.-C. *et al.* High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* **82,** 378–389 (2003).
8. Soderlund, C., Humphray, S., Dunham, A. & French, L. Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.* **10,** 1772–1787 (2000).
9. Schulte, D. *et al.* BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.). *BMC Genomics* **12,** 247 (2011).
10. Doležel, J. *et al.* Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann. Bot.* **82,** 17–26 (1998).
11. Paux, E. *et al.* A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* **322,** 101–104 (2008).
12. Madishetty, K., Condamine, P., Svensson, J. T., Rodriguez, E. & Close, T. J. An improved method to identify BAC clones using pooled overgos. *Nucleic Acids Res.* **35,** e5 (2007).
13. Lonardi, S. *et al.* Barcoding-free BAC pooling enables combinatorial selective sequencing of the barley gene space. preprint at http://arxiv.org/abs/1112.4438 (2011).
14. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7,** e32253 (2012).
15. Mayer, K. F. X. *et al.* Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* **23,** 1249–1263 (2011).
16. Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326,** 1112–1115 (2009).
17. Wicker, T. *et al.* A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant J.* **59,** 712–722 (2009).
18. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon. Nature* **463,** 763–768 (2010).
19. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436,** 793–800 (2005).
20. Paterson, A. H. *et al.* The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457,** 551–556 (2009).
21. Kronmiller, B. A. & Wise, R. P. TEnest: automated chronological annotation and visualization of nested plant transposable elements. *Plant Physiol.* **146,** 45–59 (2008).
22. Ohshima, K. & Okada, N. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet. Genome Res.* **110,** 475–490 (2005).
23. Wessler, S. R., Bureau, T. & White, S. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5,** 814–821 (1995).
24. Matsumoto, T. *et al.* Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol.* **156,** 20–28 (2011).
25. Zhang, P. *et al.* MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.* **138,** 27–37 (2005).
26. Wicker, T. *et al.* Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell* **23,** 1706–1718 (2011).
27. van der Biezen, E. A. & Jones, J. D. G. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **8,** R226–R228 (1998).
28. Wei, F., Wing, R. A. & Wise, R. P. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. *Plant Cell* **14,** 1903–1917 (2002).
29. Halterman, D. A. & Wise, R. P. A single-amino acid substitution in the sixth leucine-rich repeat of barley *MLA6* and *MLA13* alleviates dependence on *RAR1* for disease resistance signaling. *Plant J.* **38,** 215–226 (2004).

30. Seeholzer, S. *et al.* Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol. Plant Microbe Interact.* **23,** 497–509 (2010).
31. Jacobs, A. K. *et al.* An *Arabidopsis* callose synthase, GSL5, is required for wound and papillary callose formation. *Plant Cell* **15,** 2503–2513 (2003).
32. Jiao, Y. *et al.* A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nature Genet.* **41,** 258–263 (2009).
33. Conti, E. & Izaurralde, E. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17,** 316–325 (2005).
34. Kalyna, M. *et al.* Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* **40,** 2454–2469 (2012).
35. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100,** 189–192 (2003).
36. Bhuvanagiri, M., Schlitter, A. M., Hentze, M. W. & Kulozik, A. E. NMD: RNA biology meets human genetic medicine. *Biochem. J.* **430,** 365–377 (2010).
37. Rayson, S. *et al.* A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in *Arabidopsis thaliana* NMD mutants. *PLoS ONE* **7,** e31917 (2012).
38. Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C. & Riha, K. Aberrant growth and lethality of *Arabidopsis* deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Res.* **40,** 5615–5624 (2012).
39. Jeong, H.-J. *et al.* Nonsense-mediated mRNA decay factors, *UPF1* and *UPF3*, contribute to plant defense. *Plant Cell Physiol.* **52,** 2147–2156 (2011).
40. Lu, T. *et al.* Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res.* **20,** 1238–1249 (2010).
41. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482,** 339–346 (2012).
42. Chinen, M. & Tani, T. Diverse functions of nuclear non-coding RNAs in eukaryotic gene expression. *Front. Biosci.* **17,** 1402–1417 (2012).
43. Abdel-Ghani, A. H., Parzies, H. K., Omary, A. & Geiger, H. H. Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. *Theor. Appl. Genet.* **109,** 588–595 (2004).
44. Comadran, J. *et al.* Patterns of polymorphism and linkage disequilibrium in cultivated barley. *Theor. Appl. Genet.* **122,** 523–531 (2011).
45. Close, T. J. *et al.* Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* **10,** 582 (2009).
46. Thiel, T. *et al.* Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. *BMC Evol. Biol.* **9,** 209 (2009).
47. Martinez-Perez, E. & Moore, G. To check or not to check? The application of meiotic studies to plant breeding. *Curr. Opin. Plant Biol.* **11,** 222–227 (2008).
48. Halterman, D. A., Wei, F. S. & Wise, R. P. Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiol.* **131,** 558–567 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work has been supported from the following funding sources: German Ministry of Education and Research (BMBF) grant 0314000 "BARLEX" to K.F.X.M., M.P., U.S. and N.S.; Leibniz Association grant (Pakt f. Forschung und Innovation) to N.S.; European project of the 7th framework programme "TriticeaeGenome" to R.W., A.S., K.F.X.M., M.M. and N.S.; SFB F3705, of the Austrian Wissenschaftsfond (FWF) to K.F.X.M.; ERA-NET PG project "BARCODE" grant to M.M., N.S. and R.W.; Scottish Government/BBSRC grant BB/100663X/1 to R.W., D.M., P.H., J.R., M.C. and P.K.; National Science Foundation grant DBI 0321756 "Coupling EST and Bacterial Artificial Chromosome Resources to Access the Barley Genome" and DBI-1062301 "Barcoding-Free Multiplexing: Leveraging Combinatorial Pooling for High-Throughput Sequencing" to T.J.C. and S.L.; USDA-CSREES-NRI grant 2006-55606-16722 "Barley Coordinated Agricultural Project: Leveraging Genomics, Genetics, and Breeding for Gene Discovery and Barley Improvement" to G.J.M., R.P.W., T.J.C. and S.L.; the Agriculture and Food Research Initiative Plant Genome, Genetics and Breeding Program of USDA-CSREES-NIFA grant 2009-65300-05645 "Advancing the Barley Genome" to T.J.C., S.L. and G.J.M.; BRAIN and NBRP-Japan grants to K.S., Japanese MAFF Grant (TRG1008) to T.M. A full list of acknowledgements is in the Supplementary Information.

Author Contributions See list of consortium authors. R.A., D.S., H.L., B.S., S.T., M.G., F.C., T.N., M.S., M.P., H.G., P.H., T.S., K.F.X.M., R.W. and N.S. contributed equally to their respective work packages and tasks.

Author Information Sequence resources generated or compiled in this study have been deposited at EMBL/ENA or NCBI GenBank. A full list of sequence raw data accession numbers as well as URLs for data download, visualization or search are provided in Supplementary Note 1 and Supplementary Table 1. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike license, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.F.X.M (k.mayer@helmholtz-muenchen.de), R.W. (Robbie.Waugh@hutton.ac.uk) or N.S. (stein@ipk-gatersleben.de).

The International Barley Genome Sequencing Consortium (IBSC)

Principal investigators Klaus F. X. Mayer[1], Robbie Waugh[2], Peter Langridge[3], Timothy J. Close[4], Roger P. Wise[5], Andreas Graner[6], Takashi Matsumoto[7], Kazuhiro Sato[8], Alan Schulman[9], Gary J. Muehlbauer[10], Nils Stein[6]

Physical map construction and direct anchoring Ruvini Ariyadasa[6], Daniela Schulte[6], Naser Poursarebani[6], Ruonan Zhou[6], Burkhard Steuernagel[6], Martin Mascher[6], Uwe Scholz[6], Bujun Shi[3], Peter Langridge[3], Kavitha Madishetty[4], Jan T. Svensson[4], Prasanna Bhat[4], Matthew Moscou[4], Josh Resnik[4], Timothy J. Close[4], Gary J. Muehlbauer[10], Pete Hedley[2], Hui Liu[2], Jenny Morris[2], Robbie Waugh[2], Zeev Frenkel[11], Avraham Korol[11], Hélène Bergès[12], Andreas Graner[6], Nils Stein (leader)[6]

Genomic sequencing and assembly Burkhard Steuernagel[6], Uwe Scholz[6], Stefan Taudien[13], Marius Felder[13], Marco Groth[13], Matthias Platzer[13], Nils Stein (leader)[6]

BAC sequencing and assembly Burkhard Steuernagel[6], Uwe Scholz[6], Axel Himmelbach[6], Stefan Taudien[13], Marius Felder[13], Matthias Platzer[13], Stefano Lonardi[14], Denisa Duma[14], Matthew Alpert[14], Francesa Cordero[14,22], Marco Beccuti[14], Gianfranco Ciardo[14], Yaqin Ma[14], Steve Wanamaker[4], Timothy J. Close (co-leader)[4], Nils Stein (leader)[6]

BAC-end sequencing Federica Cattonaro[15], Vera Vendramin[16], Simone Scalabrin[15], Slobodanka Radovic[16], Rod Wing[17], Daniela Schulte[6], Burkhard Steuernagel[6], Michele Morgante[15,16], Nils Stein[6], Robbie Waugh (leader)[2]

Integration of physical/genetic map and sequence resources Thomas Nussbaumer[1], Heidrun Gundlach[1], Mihaela Martis[1], Ruvini Ariyadasa[6], Naser Poursarebani[6], Burkhard Steuernagel[6], Uwe Scholz[6], Roger P. Wise[5], Jesse Poland[18], Nils Stein[6], Klaus F. X. Mayer (leader)[1]

Gene annotation Manuel Spannagl[1], Matthias Pfeifer[1], Heidrun Gundlach[1], Klaus F. X. Mayer (leader)[1]

Repetitive DNA analysis Heidrun Gundlach[1], Cédric Moisy[9], Jaakko Tanskanen[9], Simone Scalabrin[15], Andrea Zuccolo[15], Vera Vendramin[16], Michele Morgante[15,16], Klaus F. X. Mayer (co-leader)[1], Alan Schulman (leader)[9]

Transcriptome sequencing and analysis Matthias Pfeifer[1], Manuel Spannagl[1], Pete Hedley[2], Jenny Morris[2], Joanne Russell[2], Arnis Druka[2], David Marshall[2], Micha Bayer[2], David Swarbreck[19], Dharanya Sampath[19], Sarah Ayling[19], Melanie Febrer[19], Mario Caccamo[19], Takashi Matsumoto[7], Tsuyoshi Tanaka[7], Kazuhiro Sato[8], Roger P. Wise[5], Timothy J. Close[4], Steve Wannamaker[4], Gary J. Muehlbauer[10], Nils Stein[6], Klaus F. X. Mayer (co-leader)[1], Robbie Waugh (leader)[2]

Re-sequencing and diversity analysis Burkhard Steuernagel[6], Thomas Schmutzer[6], Martin Mascher[6], Uwe Scholz[6], Stefan Taudien[13], Matthias Platzer[13], Kazuhiro Sato[8], David Marshall[2], Micha Bayer[2], Robbie Waugh (co-leader)[2], Nils Stein (leader)[6]

Writing and editing of the manuscript Klaus F. X. Mayer (co-leader)[1], Robbie Waugh (co-leader)[2], John W. S. Brown[2,20], Alan Schulman[9], Peter Langridge[3], Matthias Platzer[13], Geoffrey B. Fincher[21], Gary J. Muehlbauer[10], Kazuhiro Sato[8], Timothy J. Close[4], Roger P. Wise[5] & Nils Stein (leader)[6]

[1]MIPS/IBIS, Helmholtz Zentrum München, D-85764 Neuherberg, Germany. [2]The James Hutton Institute, Invergowrie, Dundee DD2 5DE, UK. [3]Australian Centre for Plant Functional Genomics, University of Adelaide, Glen Osmond 5064, Australia. [4]Department of Botany & Plant Sciences, University of California, Riverside, California 92521, USA. [5]USDA-ARS, Department of Plant Pathology & Microbiology, Iowa State University, Ames, Iowa 50011-1020, USA. [6]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Seeland OT Gatersleben, Germany. [7]National Institute of Agrobiological Sciences, 2-1-2, Kannondai, Tsukuba Ibaraki 305-8602, Japan. [8]Okayama University, Kurashiki 710-0046, Japan. [9]MTT Agrifood Research and Institute of Biotechnology, University of Helsinki, FIN-00014 Helsinki, Finland. [10]University of Minnesota, Department of Agronomy and Plant Genetics, Department of Plant Biology, St Paul, Minnesota 55108, USA. [11]Institute of Evolution, University of Haifa, Haifa 31905, Israel. [12]INRA-CNRGV, Auzeville CS 52627, France. [13]Leibniz Institute of Age Research- Fritz Lipmann Institute (FLI), D-07745 Jena, Germany. [14]Department of Computer Science & Engineering, University of California, Riverside, California 92521, USA. [15]Istituto di Genomica Applicata, Via J. Linussio 51, 33100 Udine, Italy. [16]Dipartimento di Scienze Agrarie ed Ambientali, Università di Udine, 33100 Udine, Italy. [17]University of Arizona, Arizona Genomics Institute, Tucson, Arizona 85721, USA. [18]USDA-ARS Hard Winter Wheat Genetics Research Unit and Kansas State University, Manhattan, Kansas 66506, USA. [19]The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK. [20]Division of Plant Sciences, University of Dundee at The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK. [21]ARC Centre of Excellence in Plant Cell Walls, University of Adelaide, Waite Campus, Glen Osmond, South Australia 5064, Australia. [22]Department of Computer Science, Corso Svizzera 185, 10149 Torino, Italy.

doi:10.1038/nature11605

MR1 presents microbial vitamin B metabolites to MAIT cells

Lars Kjer-Nielsen[1], Onisha Patel[2], Alexandra J. Corbett[1], Jérôme Le Nours[2,3], Bronwyn Meehan[1], Ligong Liu[4], Mugdha Bhati[2], Zhenjun Chen[1], Lyudmila Kostenko[1], Rangsima Reantragoon[1], Nicholas A. Williamson[5], Anthony W. Purcell[2,5], Nadine L. Dudek[2,5], Malcolm J. McConville[5], Richard A. J. O'Hair[6], George N. Khairallah[6], Dale I. Godfrey[1], David P. Fairlie[4], Jamie Rossjohn[2,3,7]* & James McCluskey[1]*

Antigen-presenting molecules, encoded by the major histocompatibility complex (MHC) and CD1 family, bind peptide- and lipid-based antigens, respectively, for recognition by T cells. Mucosal-associated invariant T (MAIT) cells are an abundant population of innate-like T cells in humans that are activated by an antigen(s) bound to the MHC class I-like molecule MR1. Although the identity of MR1-restricted antigen(s) is unknown, it is present in numerous bacteria and yeast. Here we show that the structure and chemistry within the antigen-binding cleft of MR1 is distinct from the MHC and CD1 families. MR1 is ideally suited to bind ligands originating from vitamin metabolites. The structure of MR1 in complex with 6-formyl pterin, a folic acid (vitamin B9) metabolite, shows the pterin ring sequestered within MR1. Furthermore, we characterize related MR1-restricted vitamin derivatives, originating from the bacterial riboflavin (vitamin B2) biosynthetic pathway, which specifically and potently activate MAIT cells. Accordingly, we show that metabolites of vitamin B represent a class of antigen that are presented by MR1 for MAIT-cell immunosurveillance. As many vitamin biosynthetic pathways are unique to bacteria and yeast, our data suggest that MAIT cells use these metabolites to detect microbial infection.

MAIT cells, which comprise up to 10% of the peripheral blood T-cell population of humans, are readily detected in blood, mesenteric lymph nodes and the gastrointestinal mucosa[1]. Despite the abundance of this innate-like T-cell population, the role of MAIT cells in health and disease is unclear, although they can have crucial functions in protective immunity[2,3]. Indeed, MAIT cells are rapidly activated by a wide range of microorganisms, including diverse strains of bacteria and yeast, suggesting that MAIT cells respond to a conserved antigen(s) common to these microbes[4–6]. MAIT cells are activated by means of their αβ T-cell antigen receptor (TCR), which, analogous to the natural killer T (NKT)-cell antigen receptor[7,8], consists of an invariant TCR α chain paired with a limited array of Vβ chains (Vα7.2Jα33 paired with Vβ2 or Vβ13)[9,10]. The constrained gene usage of the MAIT-cell antigen receptor (MAIT TCR), which is distinct from that of the NKT-cell antigen receptor (Vα24Jα18-Vβ11 in humans), suggests that MAIT cells target a key, albeit limited and atypical, class of antigen[6]. Providing further evidence for highly conserved MAIT ligand(s), mutagenesis studies of MAIT TCRs with different Vβ segments showed that a cluster of conserved TCR residues is crucial for MAIT TCR recognition of diverse microbes[9]. The MAIT TCR is restricted to the monomorphic MHC class I (MHC-I)-like related molecule (MR1), which is highly conserved in mammals and encoded by a single gene that is not associated with the MHC[11–13]. MR1 transcripts are ubiquitously expressed in all cell types, and the MR1 primary structure is highly conserved across species, thereby suggesting a key, evolutionarily conserved function for MR1 in immunity[13,14]. Endogenous cell surface expression levels of MR1 are typically very low, suggesting that quantitatively limiting antigen(s) may be required to increase MR1 presentation[12,15]. MR1 shares considerable sequence similarity with the MHC-I and CD1 families—antigen-presenting molecules that are ideally suited to bind peptide- and lipid-based antigens, respectively. However, the precise identity of the MR1-restricted antigen(s), representing an important question in MAIT-cell biology, is unknown.

Identification of an MR1-restricted ligand

Yeast and many, but not all, bacterial species can activate MAIT cells in an MR1-restricted manner, suggesting that a closely related series of antigen(s) common to these microbes specifically bind MR1 (ref. 3). MHC-I and MHC-I-like molecules are usually stable only in the presence of ligand, and so, unsurprisingly, MR1 could not be refolded efficiently in refolding buffer alone, consistent with MR1 requiring a specific antigen(s) for its assembly and stability (not shown). Reasoning that recovery of properly assembled MR1–β2-microglobulin (β2m) complexes would indicate efficient capture of a ligand(s), we established refolds of denatured human MR1 and β2m with different sources of candidate ligands, including lipid-based ligands that were previously proposed to bind MR1 (ref. 16). Refolding in the cell culture medium RPMI-1640 (a control) enhanced the yield of MR1–β2m complexes, suggesting the presence of trace amounts of MR1 ligand(s) in this medium. RPMI-1640 is a defined culture medium containing several vitamin supplements, some of which are uniquely synthesized in bacteria and plants, but not mammals. Hence, we tested the refolding of MR1 in the presence of various

---

[1]Department of Microbiology & Immunology, University of Melbourne, Parkville, Victoria 3010, Australia. [2]Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, Australia. [3]Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, Australia. [4]Division of Chemistry & Structural Biology, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. [5]Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, Metabolomics Australia, University of Melbourne, Parkville, Victoria 3010, Australia. [6]School of Chemistry, Bio21 Molecular Science and Biotechnology Institute and ARC Centre of Excellence for Free Radical Chemistry and Biotechnology, University of Melbourne, Melbourne, Victoria 3010, Australia. [7]Institute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, UK.

*These authors contributed equally to this work.

vitamin sources. We observed that folic acid sources (including RPMI-1640 medium, vitamin B complex tablets, or folic acid itself) all notably enhanced yields of MR1 purified by subsequent chromatographic steps (Fig. 1a). The chromatographic properties of the refolded MR1–antigen closely matched those of refolded peptide-MHC-I (ref. 17), with a molecular mass of approximately 44 kDa, consistent with a monomeric MR1–β2m complex (Fig. 1a). Moreover, the refolded material reacted with an anti-MR1 monoclonal antibody (Fig. 1b), indicating that a vitamin-derived compound(s) enabled the proper refolding of MR1.

Analysis of MR1 refolded with folic acid by negative mode electrospray ionization–time-of-flight mass spectrometry (ESI–TOF-MS), exclusively revealed a ligand with a mass to charge (*m*/*z*) ratio of 190.03. Comparison of this species with the Scripps Metlin metabolite mass spectrometry database (http://metlin.scripps.edu/), as well as further tandem mass spectrometry analysis, suggested that this species was 6-formyl pterin (6-FP) (Fig. 1c, d). ESI–TOF-MS analysis of MR1 refolded with pure, chemically synthesized 6-FP showed an identical species to that found in MR1 refolded with folic acid (Fig. 1d). The species giving *m*/*z* 190.03 by ESI–TOF analysis of 6-FP and MR1–6-FP had identical liquid chromatography retention times and product ions after tandem mass spectrometry analysis (Fig. 1d and Supplementary Fig. 1). Notably, 6-FP is produced by the photodegradation of folic acid[18], and folic acid contained trace amounts of 6-FP as measured by ESI–TOF-MS. Moreover, chromatographically purified folic acid that was depleted of 6-FP, and subsequently promptly used for refolding with MR1, still yielded the MR1–6-FP complex (data not shown), indicating that photodegradation of folic acid was rapid. Furthermore, 6-FP specifically upregulated cell surface expression of MR1 on human lymphoid C1R cells (Supplementary Fig. 2). These studies provided evidence that MR1 can bind to small molecule organic compounds, such as 6-FP, that are derived from vitamin metabolism.

The structure of MR1–antigen

To gain insight into the detailed architecture of the MR1 molecule, we expressed and refolded MR1 in complex with 6-FP and subsequently

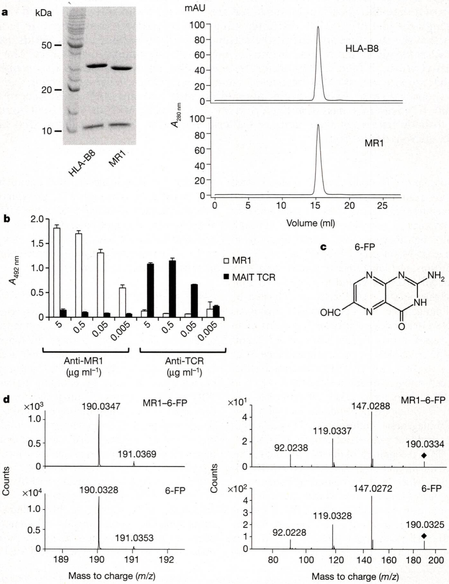


Figure 1 | Refolding of MR1 in the presence of 6-FP. a, Left, refolded MR1 separated by 15% SDS–polyacrylamide gel electrophoresis. Right, size-fractionated MR1 (bottom) and control HLA-B8 (top). mAU, milli-absorbance units. **b**, ELISA with refolded MR1 or control MAIT TCR. Data are mean and s.d. from triplicate samples ($n = 3$). **c**, 6-FP. **d**, Mass spectrometry analysis of MR1–6-FP. Left, compound at *m*/*z* 190.0347 observed with MR1–6-FP (top) and 6-FP (190.0328) (bottom). Right, product ions obtained from targeted fragmentation of the *m*/*z* 190.0347 species observed with MR1–6-FP (top) and of the *m*/*z* 190.0328 species of 6-FP (bottom); precursor ions indicated by black diamonds. Background counts were obtained with control buffer-only samples.

determined the structure of the MR1–6-FP complex (Supplementary Table 1 and Supplementary Fig. 3) and compared it with representative peptide- and lipid-binding antigen-presenting molecules, namely HLA-A2 and CD1d, respectively.

MR1 adopted a standard MHC-I fold (Fig. 2a). The MR1 heavy chain shares close structural homology with HLA-A2 (39% sequence identity, root mean square deviation (r.m.s.d.) of 1.73 Å over 213 Cα atoms) and CD1d (22% sequence identity, r.m.s.d. 2.26 Å over 151 Cα atoms)[19,20]. MR1 was most closely related to an avian monomorphic MHC-I like molecule (43% sequence identity, r.m.s.d. 1.77 Å over 276 Cα atoms) that is thought to bind non-peptide-based antigens[21] (Supplementary Fig. 4). The α1-α2 domains formed the MR1 antigen-binding cleft, which comprises two long α-helices sitting atop a β-sheet, akin to HLA-A2 and CD1d (Fig. 2b). The helices of the MR1 antigen-binding cleft were not closely juxtaposed, as was observed for the human hemochromatosis (HFE) protein, an MHC-I-like molecule that does not bind antigen[22] (Fig. 2b). Indeed, the positioning of the α1 and α2 helices of MR1 more closely resembled HLA-A2 than that of the more constricted CD1d antigen-binding cleft (r.m.s.d. 1.0 Å over 133 Cα atoms and 3.1 Å over 99 Cα atoms, respectively) (Fig. 2b). However, the central cleft of MR1 is not suited, either chemically or structurally, to accommodate peptide- or lipid-based antigens (Fig. 3a). Namely, the HLA-A2 cleft is solvent exposed and mostly polar, thereby ideally suited for binding peptides[23]; whereas the CD1d cleft binds lipids by means of a hydrophobic-lined cavity shielded from solvent[20]. By comparison, the MR1 antigen-binding cleft was mostly solvent exposed, consisting of a mixture of charged and hydrophobic residues, of which a preponderance of aromatic residues within the α1 and α2 helices was evident (Fig. 3a and Supplementary Fig. 5). Furthermore, the central cavity of MR1 (760 Å^3) was much smaller than that of CD1d (1,690 Å^3)[20] (Fig. 3b). Although MHC-I molecules contain a conserved network of residues at the amino- and carboxy-terminal ends of the antigen-binding groove that tethers the termini of antigenic peptides[23], the corresponding locations within the MR1 cleft, although showing some conservation, were different from MHC-I (Fig. 3c). Furthermore, the end of the MR1 groove is not 'open', as observed for MHC class II molecules[24]. Whereas MHC-I comprises six pockets that accommodate the side chains of the peptide, these are not present in MR1 (ref. 25). Instead, a large number of bulky side chains occupied the entire length and breadth of the cleft, and it is this architecture that probably prevented the helical jaws of MR1 packing closely together (Fig. 3c). Accordingly, the structure and chemical properties of the MR1 antigen-binding cleft were distinct from that of peptide- and lipid-based antigen-presenting molecules.

Mode of MR1–antigen presentation

6-FP was located centrally within the MR1 cleft, positioned towards the base of the β-sheet (Figs 2a and 3d). The pterin ring lies relatively flat against the β-sheet, and in a different location to abacavir, a bicyclic compound recently found to bind HLA-B*57:01 (ref. 26) (Supplementary Fig. 6). 6-FP exhibited very limited solvent accessibility, with 317 Å^2 of the available 327 Å^2 being buried by MR1 (Fig. 3a). Binding to 6-FP was dominated by hydrophobic interactions, with Tyr 7, Tyr 62, Trp 69 and Trp 156 forming an 'aromatic cradle' that sequestered the ligand. In addition, the ligand formed van der Waals interactions with Arg 9, Arg 94, Ile 96 and Gln 153 (Fig. 3d) (Supplementary Table 2). Notably, a previous mutational study on MR1 implicated Tyr 7, Arg 9 and Arg 94 in MAIT-cell activation[11]. There was clear evidence for a covalent bond between the Lys 43$^{N\zeta}$ and the formyl group of 6-FP (Supplementary Fig. 3), indicating that, during purification and/or crystallization, Lys 43 had formed a Schiff base with the formyl group. An engineered Lys43Arg MR1 mutation failed to refold in the presence of 6-FP (not shown), highlighting the importance of the Lys 43–6-FP interaction. Adjacent to the MR1–antigen-binding pocket there were two positively charged residues (Arg 9 and Arg 94) protruding up into the cleft, suggesting a requirement for polar moieties in other potential MR1-restricted ligands. Of note, the residues in contact with 6-FP (and the two Arg residues) are conserved across MR1 from all species (Supplementary Fig. 7), suggesting that recognition of ligands within the MR1-binding pocket is highly conserved. Hence, the structure of the MR1–6-FP complex shows how MR1 is ideally suited to present small organic compounds that can originate from vitamins.

MR1-restricted MAIT activation

Although 6-FP is a ligand for MR1, it did not activate Jurkat cells transduced with a MAIT TCR (termed Jurkat.MAIT cells) (Fig. 4a) or

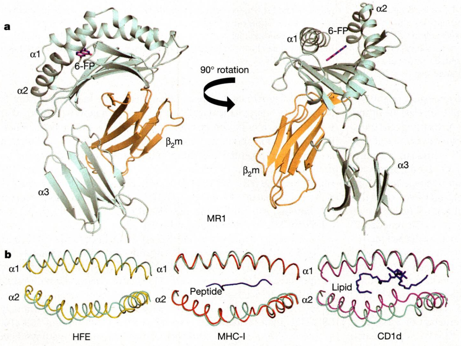


Figure 2 | Overview of the crystal structure of MR1–antigen complex. **a**, Structure of MR1. The α1, α2 and α3 domains of MR1 non-covalently attached to β$_2$m. α1, α2 and α3 are in cyan; β$_2$m is in light orange; and 6-FP is in magenta. **b**, Overlay of the α1 and α2 helices using the residues within the antigen-binding cleft of MR1 with HFE (left, PDB code 1A6Z), MHC-I (middle, HLA-A2, PDB code 3GSO) and CD1d (right, PDB code 1ZT4). MR1 is in cyan; HFE is in yellow; MHC-I is in red; CD1d is in magenta.

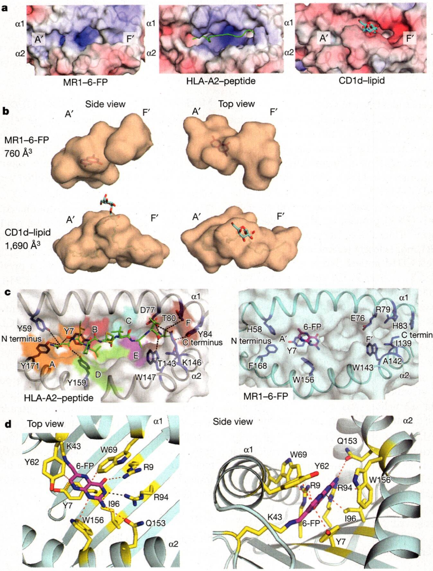

a
α1
α2
A′
F′
MR1–6-FP
α1
α2
HLA-A2–peptide
α1
α2
A′
F′
CD1d–lipid
b
Side view
Top view
A′
F′
A′
F′
MR1–6-FP
760 Å³
CD1d–lipid
1,690 Å³
A′
F′
A′
F′
c
α1
D77
T80
Y59
N terminus
Y7
B
C
F
Y84
C terminus
E
T143
K146
Y171
A
D
Y159
W147
α2
HLA-A2–peptide
α1
R79
E76
H83
H58
6-FP
N terminus
A′
Y7
C terminus
F′
I139
A142
F168
W143
W156
α2
MR1–6-FP
d
Top view
α1
K43
W69
Y62
6-FP
R9
R94
Y7
I96
W156
Q153
α2
Side view
α2
Q153
α1
W69
Y62
R9
R94
W156
K43
I96
6-FP
Y7

Figure 3 | Comparison of MR1, MHC-I and MHC-I-like binding clefts. **a**, Electrostatic surface of MR1, HLA-A2 and CD1d antigen-binding clefts. Peptide, green; lipid, cyan. A′ and F′ are binding pockets within CD1d. **b**, Side (left) and top (right) views of MR1–6-FP and CD1d–α-galactosylceramide. **c**, Left, peptide interaction with HLA-A2 residues; right, corresponding MR1 residues. Six pockets (A–F) of MHC-I, coloured in orange, red, cyan, green, magenta and brown, respectively; peptide is in green. The surface in **b** and **c** is transparent to show 6-FP binding. **d**, Top (left) and side (right) views of MR1–6-FP. MR1 residues that contact 6-FP are in yellow. H-bond and van der Waals interactions are shown in black- and red-dashed lines, respectively. MR1 is in cyan; 6-FP is in magenta.

primary MAIT cells (Fig. 5c, e), suggesting pterin analogues might provide a basic structural scaffold for another ligand, or class of ligand, capable of activating MAIT cells. Notably, antigen-presenting cells transfected with MR1 and infected with *Salmonella typhimurium* specifically activated Jurkat.MAIT cells[9], but not Jurkat cells expressing a control MHC-restricted TCR (Supplementary Fig. 8a, b). Furthermore, we found that the supernatant from *S. typhimurium* was able to activate Jurkat.MAIT and MAIT cells (Figs 4a and 5a–c, e), and we reasoned that identification of the bacterial-activating ligand(s) from *S. typhimurium* supernatant might be facilitated, if MR1–β2m could be refolded in supernatant from *Salmonella* grown in M9 minimal media lacking vitamin supplements and their derivatives, which might compete for MR1 binding. Therefore we refolded MR1 and searched for ligands that complexed with MR1 only in supernatant from *Salmonella* grown in minimal medium. Using this approach, a single compound at *m*/*z* 329.1100 was identified by high-resolution ESI–TOF-MS. This was exclusively present in MR1 refolded in the presence of supernatant of *Salmonella* grown in M9 minimal media (that is, lacking vitamin supplements) (Fig. 4b, c). Other MR1 ligands specifically derived from the *Salmonella* supernatant were not identified in the mass range from 50 to 1,000 AMU (Supplementary Fig. 9). This compound was assigned an unambiguous atomic composition of $C_{12}H_{18}N_4O_7$ on the basis of both its molecular ion and its component isotopic mass distribution pattern. A search for potential matching compounds suggested one of the derivatives of riboflavin (vitamin B2); reduced 6-hydroxymethyl-8-D-ribityllumazine (rRL-6-CH$_2$OH) ($C_{12}H_{18}N_4O_7$), 7-hydroxy-6-methyl-8-D-ribityllumazine (RL-6-Me-7-OH) ($C_{12}H_{16}N_4O_7$) and its precursor, 6,7-dimethyl-8-D-ribityllumazine (RL-6,7-diMe) ($C_{13}H_{16}N_4O_6$) are candidate MAIT-activating ligands (Fig. 4d).

To establish formally whether rRL-6-CH$_2$OH, RL-6-Me-7-OH and/or RL-6,7-diMe represented MR1 ligands and could activate MAIT cells, we chemically synthesized and biochemically characterized these compounds. Analysis by ESI–TOF-MS of pure, chemically synthesized rRL-6-CH$_2$OH exclusively revealed a species identical to that found in MR1 refolded with *Salmonella* supernatant (Fig. 4b, c). Namely, ESI–TOF analysis identified a species with an *m*/*z* ratio of 329.1100 in the MR1 *Salmonella* supernatant that matched precisely with synthetic rRL-6-CH$_2$OH (found 329.1116, calculated for $C_{12}H_{17}N_4O_7^-$ 329.1103,

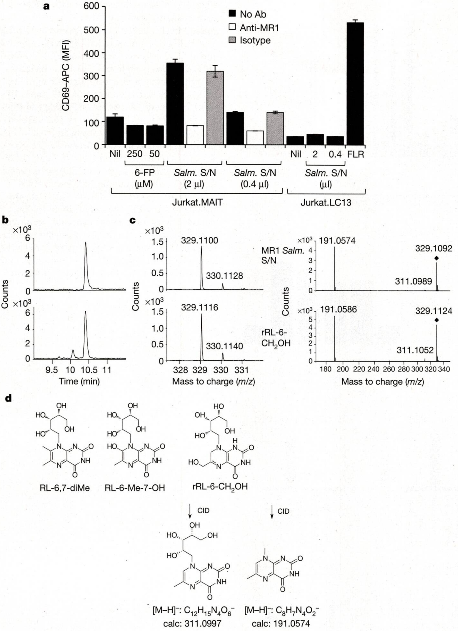


Figure 4 | Identification of bacterially-derived MAIT-cell antigens. **a**, CD69 expression of Jurkat.MAIT and control Jurkat.LC13 cells with 6-FP or *Salmonella* supernatant (*Salm.* S/N), and C1R.MR1 cells; blocking by an anti-MR1 monoclonal antibody. APC, allophycocyanin; MFI, mean fluorescent intensity. Activation of control Jurkat.LC13 cells by FLRGRAYGL (FLR) peptide with C1R.HLA-B8 cells. Data are mean and s.e.m. from triplicate samples ($n = 3$). **b**, Extracted ion chromatograms (EIC) showing retention times of the m/z 329.1000 species from MR1 refolded with *Salmonella* supernatant (top) and the m/z 329.1116 species from synthetic rRL-6-CH_2OH (bottom). **c**, Left, compound with m/z 329.1100 from MR1 and *Salmonella* supernatant (top), and m/z 329.1116 for synthetic rRL-6-CH_2OH (bottom). Right, product ions from targeted fragmentation of MR1 and *Salmonella* supernatant and of rRL-6-CH_2OH (precursor ions indicated by black diamonds). **d**, RL-6,7-diMe, RL-6-Me-7-OH, rRL-6-CH_2OH and its product ions. CID, collision-induced dissociation.

$[M-H]^-$). Moreover, they showed identical isotopic mass distribution patterns (confirming atomic composition), identical fragmentation patterns by tandem mass spectrometry analysis, and had identical liquid chromatography–mass spectrometry (LC–MS) column retention times (Fig. 4b–d). The structures of tandem mass spectrometry fragment ions provide important evidence for the assigned structure of rRL-6-CH_2OH, which undergoes dehydration/tautomerism to give m/z 311.1052 (6-methyl-8-D-ribityllumazine) or sequential dehydration/ribityl side-chain scission to give m/z 191.0586 (Fig. 4c, d). The structures of rRL-6-CH_2OH, RL-6-Me-7-OH and RL-6,7-diMe are closely related to 6-FP, but possess an extra ribityl moiety that, based on the crystal structure of MR1–6-FP, may permit direct contact by the MAIT TCR. Notably, these compounds are derived from the riboflavin biosynthetic pathway present in most, but not all, bacteria and yeast (Supplementary Table 3).

We tested the ability of these compounds to activate Jurkat.MAIT cells and human MAIT cells from peripheral blood (Fig. 5). Although the ribityl lumazines failed to activate the control Jurkat.LC13 cell line in the presence of MR1-expressing C1R cells (Fig. 5a), they specifically activated three Jurkat.MAIT-cell lines (transduced with TRBV6.1, TRBV6.4 and TRBV20 MAIT TCRs) in the presence of C1R cells expressing MR1 (Fig. 5a), whereas riboflavin did not (not shown). An anti-MR1 blocking monoclonal antibody[11] specifically inhibited the riboflavin metabolite-mediated Jurkat.MAIT activation (Fig. 5b and Supplementary Fig. 10), and rRL-6-CH_2OH upregulated MR1 cell surface expression on C1R cells (Supplementary Fig. 2). All compounds specifically activated freshly isolated MAIT cells as defined by $CD3^+$ $CD4^-$ $CD161^+$ $TRAV1.2^+$ (monoclonal antibody $D5^+$), but not other $D5^-$ $CD161^-$ or $D5^+$ $CD161^-$ $CD3^+$ cells (Fig. 5c–e and Supplementary Figs 11 and 12). Activation was assayed by CD69 upregulation (Fig. 5c, d) and intracellular cytokine staining for interferon (IFN)-γ and tumour necrosis factor (TNF) (Fig. 5e). Notably, rRL-6-CH_2OH, although closely related to RL-6-Me-7-OH and RL-6,7-diMe, was a much more potent MAIT agonist as judged by Jurkat.MAIT-cell and MAIT-cell activation. This indicates that there are several riboflavin-based metabolites that exhibit a broad spectrum in their ability to activate MAIT cells. The enzymatic pathway that generates these riboflavin precursors only seems to be found in

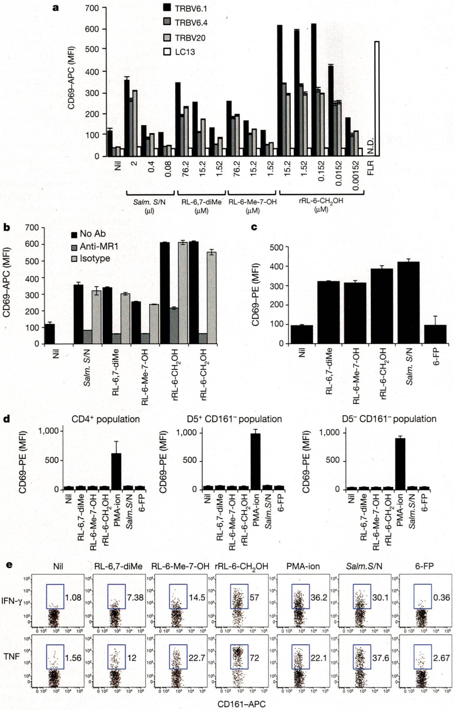


Figure 5 | MAIT-cell activation with MR1-restricted antigens. **a**, **b**, Activation of Jurkat.MAIT cells (TRBV6.1, TRBV6.4 or TRBV20 TCRs), but not control Jurkat.LC13 cells. Control Jurkat.LC13 cells were activated by FLRGRAYGL (FLR; 285 nM) peptide and C1R.HLA-B8 cells (**a**, right). In the same experiment, Jurkat.MAIT TRBV6.1 cell activation was blocked by an MR1-specific antibody (Ab; **b**). (RL-6,7-diMe and RL-6-Me-7-OH: 76 µM; rRL-6-CH_2OH: 15.2 µM (left), 0.152 µM (right).) Data are mean and s.e.m. N.D., not determined. **c**, **d**, MAIT activation (RL-6,7-diMe, RL-6-Me-7-OH and 6-FP: 76.2 µM; rRL-6-CH_2OH: 0.152 µM) assayed by CD69 expression for MAIT cells (**c**) and non-MAIT cells ($CD4^+$; $D5^+$ $CD161^-$; and $D5^-$ $CD161^-$) (**d**). Data are mean and s.e.m. PE, phycoerythrin; PMA-ion, phorbol myristate acetate (PMA) and ionomycin. **e**, MAIT-cell activation as assayed by intracellular staining for IFN-γ and TNF on gated MAIT cells. Compound concentrations are as in **c** and **d**. One representative sample from triplicates is shown.

microbes that are capable of activating MAIT cells and is absent in non-activating microbes (Supplementary Table 3). Accordingly, direct precursors and metabolites of riboflavin biosynthesis clearly activate MAIT cells.

Discussion

T cells expressing αβ TCRs interact with diverse antigens presented by MHC-I and MHC-I-like molecules[27]. Indeed, the cellular arm of the immune system uses an array of polymorphic and monomorphic antigen-presenting molecules to provide collectively broad awareness of signature microbial products with distinct chemistries, thereby enabling self from non-self discrimination and facilitating protective immunity. Our findings provide evidence that this antigenic arsenal is extended by the capacity of MR1 to present small organic compounds derived from biosynthetic intermediates to, or chemical metabolites of, endogenously synthesized vitamins.

MAIT cells are activated, in an MR1-dependent manner, by a broad spectrum of bacteria and yeast, but not by viruses or certain bacteria[3]. Although a folic acid (vitamin B9) derivative (6-FP) could bind MR1, it was non-stimulatory for MAIT cells, suggesting that this ligand represents a structural scaffold for binding MR1, but requires a further moiety to engender MAIT-cell activation. Accordingly, we showed that riboflavin (vitamin B2) derivatives, larger in size but closely related structurally to 6-FP, could bind MR1 and activate MAIT cells. A key distinction between the microbes that are stimulatory and non-stimulatory to MAIT cells is that the former synthesize riboflavin, whereas the latter do not. These observations suggest a mode by which MAIT cells might sense microbial infection or overgrowth at mucosal sites in an MR1-restricted manner. Of note, these vitamin-based metabolites are secreted and are diffusible, suggesting a mechanism by which MAIT cells might sense bacterial activity across mucosal membranes potentially regulating local immunity and mucosal barrier functions. Our findings highlight that bacterially produced vitamin-based metabolites can exhibit a broad spectrum of MAIT-cell activation. Indeed, the identification of both activating and non-activating MR1 ligands suggests competition between these agents thus modulating MAIT-cell activity in the context of bacterial infection. Moreover, as the pterin ring occurs widely in nature, and represents a common scaffold of small molecule therapeutics, it will be of interest to establish whether the MAIT cell–MR1 axis is perturbed in any pathological or drug-induced conditions; how diet can influence MAIT activity; and whether other microbial metabolites represent MAIT-cell ligands capable of modulating the function of these cells. Clearly, there is a close relationship between the human immune system, host fitness and gut microbiota metabolites[28–30]. Defining MAIT-cell-activating ligands represents a fundamental new advance that will be pivotal in understanding the physiological and pathological role of MAIT cells, a very abundant population of innate-like T cells associated with the gastrointestinal mucosa. Our findings also suggest that other microbial-specific metabolites[30] may serve as molecular signatures of microbial infection that undergo immunosurveillance.

METHODS SUMMARY

Refolding MR1 with 6-FP. This was performed using a method similar to that established for MHC-I refolding. Synthesized 6-FP and 6,7-dimethyl-pterin were purchased from Schirks Laboratories, folic acid was purchased from Sigma-Aldrich. Further details are provided in the Methods.

Analysis of MR1–6-FP by mass spectrometry. MR1–6-FP (4 μg) was loaded onto an XBridge C18 reversed phase column (Waters) in 20 mM ammonium acetate, pH 5.4, buffer, and detected in an Agilent ESI-TOF mass spectrometer after elution in an acetonitrile gradient. Data was collected in negative ion mode. Synthetic 6-FP (0.4 μg) was analysed under the same conditions.

Crystallization and data collection. The MR1–6-FP complex was crystallized and the structure determined as described in the Methods.

Analysis of MR1-bacterial supernatant by mass spectrometry. See Methods for full details.

Synthesis of MR1 antigens. Chemical synthesis of RL-6,7-diMe, RL-6-Me-7-OH and reduced rRL-6-CH_2OH is described in Methods and Supplementary Fig. 13.

Activation of Jurkat.MAIT and MAIT cells. The activation of Jurkat.MAIT cells was performed as described[9]. Further details are provided in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 14 June; accepted 21 September 2012.

Published online 10 October 2012.

1. Treiner, E. *et al.* Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature* **422,** 164–169 (2003).
2. Gold, M. C. *et al.* Human mucosal associated invariant T cells detect bacterially infected cells. *PLoS Biol.* **8,** e1000407 (2010).
3. Le Bourhis, L. *et al.* Antimicrobial activity of mucosal-associated invariant T cells. *Nature Immunol.* **11,** 701–708 (2010).
4. Gapin, L. Where do MAIT cells fit in the family of unconventional T cells? *PLoS Biol.* **7,** e1000070 (2009).
5. Le Bourhis, L. *et al.* Mucosal-associated invariant T cells: unconventional development and function. *Trends Immunol.* **32,** 212–218 (2011).
6. Godfrey, D. I., Rossjohn, J. & McCluskey, J. Fighting infection with your MAITs. *Nature Immunol.* **11,** 693–695 (2010).
7. Bendelac, A., Savage, P. B. & Teyton, L. The biology of NKT cells. *Annu. Rev. Immunol.* **25,** 297–336 (2007).
8. Godfrey, D. I. *et al.* Antigen recognition by CD1d-restricted NKT T cell receptors. *Semin. Immunol.* **22,** 61–67 (2010).
9. Reantragoon, R. *et al.* Structural insight into MR1-mediated recognition of the mucosal associated invariant T cell receptor. *J. Exp. Med.* **209,** 761–774 (2012).
10. Tilloy, F. *et al.* An invariant T cell receptor α chain defines a novel TAP-independent major histocompatibility complex class Ib-restricted α/β T cell subpopulation in mammals. *J. Exp. Med.* **189,** 1907–1921 (1999).
11. Huang, S. *et al.* Evidence for MR1 antigen presentation to mucosal-associated invariant T cells. *J. Biol. Chem.* **280,** 21183–21193 (2005).
12. Huang, S. *et al.* MR1 uses an endocytic pathway to activate mucosal-associated invariant T cells. *J. Exp. Med.* **205,** 1201–1211 (2008).
13. Huang, S. *et al.* MR1 antigen presentation to mucosal-associated invariant T cells was highly conserved in evolution. *Proc. Natl Acad. Sci. USA* **106,** 8290–8295 (2009).
14. Goldfinch, N. *et al.* Conservation of mucosal associated invariant T (MAIT) cells and the MR1 restriction element in ruminants, and abundance of MAIT cells in spleen. *Vet. Res.* **41,** 62 (2010).
15. Chua, W.-J. *et al.* Endogenous MHC-related protein 1 is transiently expressed on the plasma membrane in a conformation that activates mucosal-associated invariant T cells. *J. Immunol.* **186,** 4744–4750 (2011).
16. Shimamura, M. *et al.* Modulation of Vα19 NKT cell immune responses by α-mannosyl ceramide derivatives consisting of a series of modified sphingosines. *Eur. J. Immunol.* **37,** 1836–1844 (2007).
17. Kjer-Nielsen, L. *et al.* The structure of HLA-B8 complexed to an immunodominant viral determinant: peptide-induced conformational changes and a mode of MHC class I dimerization. *J. Immunol.* **169,** 5153–5160 (2002).
18. Off, M. K. *et al.* Ultraviolet photodegradation of folic acid. *J. Photochem. Photobiol. B* **80,** 47–55 (2005).
19. Gras, S. *et al.* Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. *J. Immunol.* **183,** 430–437 (2009).
20. Koch, M. *et al.* The crystal structure of human CD1d with and without α-galactosylceramide. *Nature Immunol.* **6,** 819–826 (2005).
21. Hee, C. S. *et al.* Structure of a classical MHC class I molecule that binds "non-classical" ligands. *PLoS Biol.* **8,** e1000557 (2010).
22. Lebrón, J. A. *et al.* Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. *Cell* **93,** 111–123 (1998).
23. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329,** 506–512 (1987).
24. Stern, L. J. *et al.* Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368,** 215–221 (1994).
25. Garrett, T. P., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* **342,** 692–696 (1989).
26. Illing, P. T. *et al.* Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature* **486,** 554–558 (2012).
27. Godfrey, D. I., Rossjohn, J. & McCluskey, J. The fidelity, occasional promiscuity, and versatility of T cell receptor recognition. *Immunity* **28,** 304–314 (2008).
28. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336,** 1268–1273 (2012).
29. The Human Microbiome Project Consortium.. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
30. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336,** 1262–1267 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Strugnell, T. Stinear, T. Mulhern, P. O'Donnell, J. Pyke, T. Rupasinghe, D. L. Tull, J. Ralton, L. Foster, S. H. Ramarathinam, M. Bharadwaj, D. Pellicci and K. Wun for discussions and technical advice, T. Hansen for the anti-MR1 monoclonal antibody and the staff of the Australian Synchrotron for assistance with data collection. This research was supported by the National Health and Medical Research Council of Australia (NHMRC) and the Australian Research Council. O.P. was supported by an ARC Future Fellowship; A.W.P. by an NHMRC Senior Research Fellowship; M.J.M. by a NHMRC Principal Research Fellowship; D.I.G. and D.P.F. were supported by NHMRC Senior Principal Research Fellowships; J.R. was supported by an NHMRC Australia Fellowship.

Author Contributions L.K.-N. identified the MR1 and MAIT ligands, undertook analysis, performed experiments and contributed to manuscript preparation. O.P. and J.L.N. solved the structure of MR1, conducted analyses and contributed to manuscript preparation. B.M., A.J.C., M.B., A.J.C., L.K., R.R., N.A.W., A.W.P., N.L.D., M.J.M., R.A.J.O.'H., G.N.K. and D.I.G. performed experiments and/or analysed data and/or provided intellectual input or helped to write the manuscript. L.L. and D.P.F. synthesized and devised the MAIT-cell activating ligands and contributed to writing the manuscript. J.M. and J.R. co-led the investigation and contributed to design and interpretation of data, project management, and writing of the manuscript.

Author Information The atomic coordinates and structure factors for the MR1–antigen complex were deposited in the Protein Data Bank (PDB) under accession code 4GUP. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. (jamesm1@unimelb.edu.au) or J.R. (jamie.rossjohn@monash.edu).

ARTICLE

doi:10.1038/nature11498

Visualizing transient low-populated structures of RNA

Elizabeth A. Dethoff[1]*, Katja Petzold[1]*, Jeetender Chugh[1]*, Anette Casiano-Negroni[1]† & Hashim M. Al-Hashimi[1]

The visualization of RNA conformational changes has provided fundamental insights into how regulatory RNAs carry out their biological functions. The RNA structural transitions that have been characterized so far involve long-lived species that can be captured by structure characterization techniques. Here we report the nuclear magnetic resonance visualization of RNA transitions towards 'invisible' excited states (ESs), which exist in too little abundance (2–13%) and for too short a duration (45–250 µs) to allow structural characterization by conventional techniques. Transitions towards ESs result in localized rearrangements in base-pairing that alter building block elements of RNA architecture, including helix-junction-helix motifs and apical loops. The ES can inhibit function by sequestering residues involved in recognition and signalling or promote ATP-independent strand exchange. Thus, RNAs do not adopt a single conformation, but rather exist in rapid equilibrium with alternative ESs, which can be stabilized by cellular cues to affect functional outcomes.

Nuclear magnetic resonance (NMR) relaxation dispersion methods[1,2], which measure microsecond-to-millisecond conformational exchange, have made it possible to characterize the transient, low-populated excited state (ES) structures of proteins[2,3] and to establish their importance in catalysis[4], folding[5,6], signalling[7] and recognition[8]. These ESs exist in too little abundance (typically with populations $<5\%$) and for too short a duration (lifetime $<$ milliseconds) to allow structural characterization by conventional techniques. Recent advances that extend the timescale sensitivity of rotating frame ($R_{1\rho}$) carbon relaxation dispersion experiments have made it possible to characterize fully exchange processes in nucleic acids[9–11], culminating in the discovery of ES Hoogsteen base pairs in DNA[12]. Although evidence for RNA ESs has been reported for decades, their structure and role in function have remained elusive[13–15].

Here we report a strategy for characterizing the ES structures of RNA that combines $R_{1\rho}$ NMR experiments, mutagenesis and secondary structure prediction. With this approach, we visualized ES structures for three distinct RNAs and obtained insights into their biological functions.

ES structure of the HIV TAR apical loop

We used a low spin-lock field $R_{1\rho}$ NMR experiment[9–11] to measure microsecond-to-millisecond conformational exchange at sugar (C1′) and nucleobase (C8 and C6) carbon sites in the well studied hexanucleotide apical loop of the transactivation response element (TAR)[16] from the human immunodeficiency virus type-1 (HIV-1). The TAR apical loop is a flexible recognition site that allows adaptive binding to a variety of proteins[17]. We observed conformational exchange (Fig. 1b and Supplementary Fig. 1) at carbon sites spread throughout the entire TAR apical loop (Fig. 1a). The $R_{1\rho}$ data could be collectively fitted to a two-state ($\mathrm{GS} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{ES}$) exchange process (where GS indicates ground state) that is directed towards an ES with population $p_{\mathrm{ES}} \approx 13\%$ and lifetime ($\tau_{\mathrm{ES}} = 1/k_{-1}$) ≈ 45 µs (Supplementary Table 1). A slower exchange process is observed at C1′ of G33 (G33-C1′), G33-C8 and A35-C8 ($p_{\mathrm{ES}} < 1\%$ and $\tau_{\mathrm{ES}} = 1.9$–2.3 ms), which can be assigned to a distinct higher energy ES that will not be discussed further (Supplementary Discussion and Supplementary Fig. 5).

In the ground state (GS), apical loop residues exist in equilibrium between C2′-endo and C3′-endo sugar puckers, G34 forms a flexible cross-loop C30•G34 Watson–Crick (WC) base pair, whereas the bases of U31, G32 and A35 are flexible[18,19]. To gain insights into the ES structure, we examined the sugar and base ES carbon chemical shifts (ω_{ES}) obtained from the two-state analysis of the $R_{1\rho}$ data, which are sensitive reporters of base stacking, sugar pucker and *syn* versus *anti* glycosidic angles[20]. The downfield-shifted sugar ES C30-C1′, U31-C1′ and A35-C1′ chemical shifts strongly suggest that in the ES these residues adopt a pure C3′-endo sugar pucker characteristic of a helical conformation (Fig. 1a and Supplementary Table 1). The downfield-shifted base ES G34-C8 can unambiguously be assigned to a *syn* base[21] (Supplementary Discussion) and has a chemical shift that is highly characteristic of a $\underline{\mathrm{U}}\mathrm{UC}\underline{G}$ tetraloop, which features a *trans*-wobble G•U base pair (underlined) and a *syn* base (italics)[22]. Notably, TAR can accommodate a similar $\underline{\mathrm{U}}_{31}\mathrm{G}_{32}\mathrm{G}_{33}\underline{\mathrm{G}}_{34}$ tetraloop. This places G34 in a *syn* position, where it can base pair with U31, thus explaining exchange at U31-C6. It also leads to the formation of C30•A35 and U31•G34 non-canonical closing base pairs, explaining the helical conformation observed for these residues in the ES. Transitions towards this ES require disruption of the cross-strand C30•G34 base pair, explaining the measured activation free energy (12.6 kcal mol^{-1}) (Supplementary Fig. 2), which is at the low end of the free energy range required to open RNA WC base pairs (13–16 kcal mol^{-1})[23]. This ES is also predicted to be the second most energetically favourable conformation using the secondary structure prediction program MC-Fold[24] (Supplementary Fig. 3).

We used a 'mutate-and-chemical-shift-fingerprint' (MCSF) strategy to test the proposed TAR ES. Here, a mutation or chemical modification is introduced to stabilize (or destabilize) a candidate ES, and the mutant's NMR carbon chemical shift fingerprints are compared with those of the ES (or GS). We stabilized the proposed TAR ES using two point mutations, C30U (TAR(C30U)) and A35G (TAR(A35G)), that replace the ES C30•A35 non-canonical base pair with more stable WC U30•A35 and C30•G35 base pairs, respectively (Fig. 1c). Both mutants adopted the proposed ES structure, as confirmed by NMR

---

[1] Department of Chemistry & Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, USA. †Present address: NYMIRUM, 3510 West Liberty Road, Ann Arbor, Michigan 48103, USA.

*These authors contributed equally to this work.

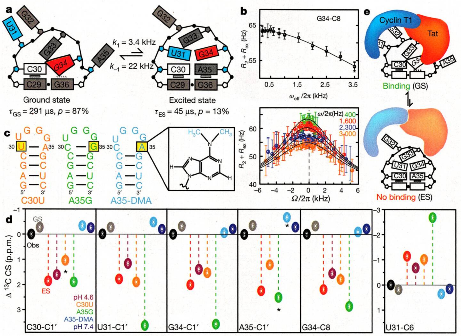


Figure 1 | Excited-state structure of the HIV-1 TAR apical loop. a, GS and ES structures of TAR. ES chemical shifts indicating increased stacking and/or *anti* glycosidic angles and C3′-endo sugar pucker are blue; decreased stacking and/or *syn* glycosidic angles and non-C3′-endo sugar pucker are red. Sites with little to no fast exchange are grey. **b**, Example relaxation dispersion profile showing dependence of $R_2 + R_{ex}$ on spin-lock power ($\omega_{eff}/2\pi$) and offset ($\Omega/2\pi$), where Ω is the difference between the observed resonance frequency and the spin-lock carrier frequency. Shown is a global fit (solid line) to a two-state Laguerre equation. Error bars indicate one s.d. **c**, Mutant mimics of GS and ES. **d**, Comparison of carbon chemical shifts (CS) for the ES, GS and mutant mimics. Carbons at the site of mutation are indicated using an asterisk. **e**, Proposed functional role for TAR ES.

(Supplementary Fig. 4), and relative to wild-type TAR both mutants featured large changes in the carbon chemical shifts, specifically at sites showing fast exchange in wild-type TAR, that are directed towards the ES chemical shifts (Fig. 1d). Inversely, we destabilized the ES by introducing a bulky N6-N6-dimethyl-substituent at the A35-N6 position (TAR(A35-DMA)) (Fig. 1c), which impairs formation of the ES C30•A35 base pair without affecting the bulged-out A35 GS conformation. This modification quenched the chemical exchange (Supplementary Fig. 1) and resulted in oppositely shifted chemical shift perturbations that are directed towards the GS (Fig. 1d). It also allowed observation of the A35-C2H2 resonance, which was otherwise severely exchange-broadened (Supplementary Fig. 4), possibly due to protonation of A35-N1 and formation of a protonated ES A35^{+}•C30 wobble base pair[15,25]. Indeed, we were able to stabilize the ES by reducing the pH from 6.4 to 4.6, as verified by analysis of carbon chemical shifts and nuclear Overhauser effects (Fig. 1d and Supplementary Figs 1 and 4). Conversely, increasing the pH to 7.4 stabilized the GS and quenched the chemical exchange (Fig. 1d and Supplementary Fig. 1).

What is the functional significance of the TAR ES? The ES sequesters U31, G34, C30 and A35 into base pairs, such that they are no longer available to bind the viral transactivator protein Tat and human cyclin T1 (Fig. 1e), which together activate transcription of the HIV-1 genome. Notably, analysis of previous mutations reveals that mutants that stabilize the TAR ES inhibit Tat/cyclin T1 binding and transcriptional activation, whereas mutants that do not stabilize the ES have little to no effect[26–28] (Supplementary Fig. 6). The TAR ES is destabilized relative to the GS by only ~1.1 kcal mol^{-1} (Supplementary Fig. 2), and can readily become >50% populated upon binding to one of several proteins known to bind TAR and interact with the apical loop[17], or by other physiochemical parameters such as the lowering of pH. The TAR ES may be involved in downregulating transactivation of the HIV genome or provide a mechanism for releasing Tat and cyclin T1. Although these functional roles remain to be verified, stabilizing the autoinhibited TAR ES immediately provides a new route for targeting TAR in the development of anti-HIV therapeutics.

ES structure of the ribosomal A-site

We used our strategy to characterize the ES structure of the ribosomal A-site internal loop[29] (Fig. 2a). The A-site has essential roles in decoding messenger RNA by flipping out two internal-loop adenines (A1492 and A1493, referred to hereafter as A92 and A93), which interact with and stabilize the codon–anticodon mini-helix formed between the cognate aminoacyl tRNA and mRNA[29,30] (see Fig. 2e). We observed extensive carbon chemical exchange at seven residues within and below the A-site internal loop (Fig. 2a, b and Supplementary Fig. 1). A two-state analysis of the $R_{1\rho}$ data revealed a global exchange process directed towards an ES with population $p_{ES} \approx 2.5\%$ and lifetime $\tau_{ES} = 1/k_{-1} \approx 248$ μs (Supplementary Table 1).

Biophysical studies show that in the GS, A92 is looped inside, probably forming a base pair with A08, whereas A93 is partially flipped out and flexible[31] (Fig. 2a). An ES involving the flipping out of A92 and A93, as observed in several X-ray and NMR structures of drug-bound A-site[32], can be ruled out based on the observation of exchange below the internal loop, ES chemical shift fingerprints that suggest increased stacking for A93 (Supplementary Fig. 5), and by comparison of ES chemical shifts with those of drug-bound A-site (Supplementary Fig. 7).

Rather, the breadth of exchange across many different residues points to a larger structural rearrangement. The downfield-shifted base ES chemical shift for U95-C6 indicates looping out of U95, whereas the upfield-shifted base carbon ES chemical shifts indicate increased stacking for A92, A93, G94 and C96 (Fig. 2a and Supplementary Table 1). These data can be explained by an alternative structure in which U95 bulges out while A93•C07, G94•U06 and A08•A92 form three consecutive non-canonical base pairs (Fig. 2d and Supplementary Table 1). A transition towards such an ES requires the opening of C07•G94, explaining the sizable free-energy barrier of ~14.8 kcal mol^{-1} (Supplementary Fig. 2)[23]. This ES is predicted by MC-Fold to be the second most energetically favourable secondary structure (Supplementary Fig. 3) and has previously been observed in molecular dynamics simulations[33].

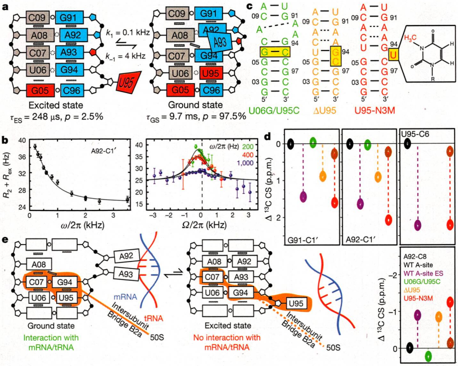


Figure 2 | Excited-state structure of the ribosomal A-site internal loop. **a**, GS and ES structures of the A-site. Chemical shift fingerprints are colour-coded as in Fig. 1a. **b**, Example relaxation dispersion profile (as in Fig. 1b). **c**, Mutant mimics of GS and ES. **d**, Comparison of carbon chemical shifts for the ES, GS and mutant mimics. **e**, Proposed functional role for A-site ES.

We confirmed the proposed A-site ES using MCSF analysis. We were able to block transitions towards the ES by replacing U06•U95 with a more stable WC G06•C95 base pair (A-site(U06G/U95C)) (Fig. 2c). This locked the A-site into the GS as judged by the GS-like chemical shifts (Fig. 2d and Supplementary Fig. 5) and absence of chemical exchange, including at sites (for example, A92 and A93) that are distant from the site of mutation (Supplementary Fig. 1). This also confirmed that all sites experience a common global exchange process. We then stabilized the proposed ES by deleting U95, which bulges out in the ES (A-site(ΔU95)), and by introducing a methyl group at U95-N3 (A-site(U95-N3M)), which is expected to disrupt the GS U06•U95 non-canonical base pair in favour of the bulged-out ES conformation (Fig. 2c). The A-site(ΔU95) mutant adopted the proposed ES structure as confirmed by NMR (Supplementary Fig. 4) and resulted in large changes in the carbon chemical shifts specifically at sites showing exchange that are directed towards the ES chemical shifts (Fig. 2d and Supplementary Fig. 5). More dramatically, the A-site(U95-N3M) mutant exhibited two equally populated sets of resonances in slow exchange on the NMR timescale (Supplementary Fig. 4), with one set corresponding to the GS and the other in near-perfect agreement with the ES (Fig. 2d and Supplementary Fig. 5).

The A-site ES sequesters A92 and A93 into base pairs, making them unavailable to decode mRNA. It also affects the structural presentation of A-site residues involved in protein recognition and formation of the B2a intersubunit crossbridge (Fig. 2a, e). Thus, we analysed previous mutational data in light of the ES A-site structure determined here. Interestingly, mutants that are predicted to stabilize the A-site ES increase the rates of stop-codon readthrough and frame-shifting, both of which are processes that can bypass mRNA decoding[34] or inhibit binding of initiation factor 1[35] (Supplementary Fig. 6). In addition, the introduction of chemical groups at the U95-N3 position, a modification that is analogous to that which we used to trap the A-site ES, leads to severely impaired association of ribosomal subunits *in vitro* due to disruption of the B2a intersubunit crossbridge[36]. This provides strong evidence that the A-site ES can form within the ribosome context where it can affect function. Although X-ray structures of the ribosome show the A-site in a GS-like conformation, in several cases, the electron density at the A-site is poor as judged by elevated B-factors, and can accommodate the ES conformation determined here (data not shown). The A-site ES invites reassessment of the A-site region in current ribosome structures and suggests a new route for targeting the A-site in the development of antibiotics.

Two ES structures in HIV-1 stem loop 1

Finally, we used our strategy to study the ES structure of the HIV-1 stem loop 1 (SL1) (Fig. 3a). SL1 spontaneously forms kissing dimers, which isomerize during viral maturation into more stable duplex dimers through mechanisms that remain poorly understood[37–39] (see Fig. 3e). This isomerization requires the melting and re-annealing of the SL1 hairpin and is catalysed *in vivo* by the nucleic acid chaperone nucleocapsid protein, but can also occur spontaneously *in vitro*[39–41]. A highly conserved asymmetric SL1 internal loop is essential for both nucleocapsid-dependent and spontaneous isomerization[42], and has been shown to induce complex NMR chemical exchange[43,44].

We observed extensive conformational exchange in a monomeric SL1 construct (SL1m)[43,45] spanning 7 base pairs in and around the internal loop (Fig. 3a, b and Supplementary Fig. 1). Unlike the A-site, the exchange extends to residues both below and above the internal loop (Fig. 3a) and cannot be globally fitted to a single process (Supplementary Table 1). Rather, at least two distinct ESs (ES1 and ES2) need to be invoked that are sensed by residues above (ES1, p_{ES1} ~9%, $\tau_{ES1} = 1/k_{-1} \approx 120$ μs) and below (ES2, $p_{ES2} \approx 2\%$, $\tau_{ES2} = 1/k_{-2}$ ~200 μs) the internal loop (Fig. 3a). Interestingly, MC-Fold also predicts a complex energy landscape for SL1m with several isoenergetic secondary structures that feature variable degrees of upward or downward migration of the bulge (Supplementary Fig. 3). This, together with the ES carbon chemical shift fingerprints and MCSF analysis, led us to deduce structures for ES1 and ES2 that feature upward and downward migration of the bulge, respectively (Supplementary Discussion).

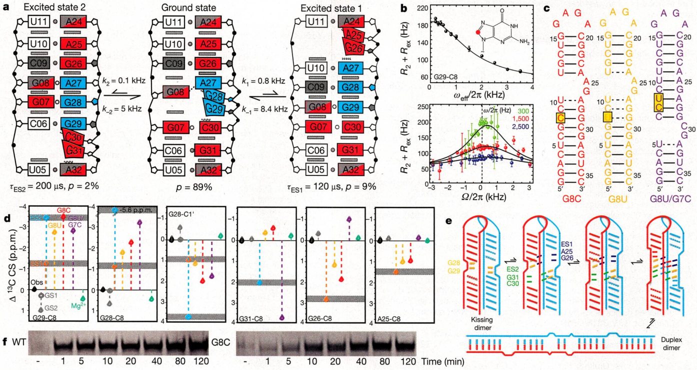


Figure 3 | Two mutually exclusive excited-state structures in HIV-1 stem-loop 1. a, GS and ES structures of SL1m. Chemical shift fingerprints are colour-coded as in Fig. 1a. **b,** Example relaxation dispersion profile (as in Fig. 1b). **c,** Mutant mimics of GS and ES. **d,** Comparison of carbon chemical shifts for the ESs, GS and mutant mimics. **e,** Proposed mechanism for spontaneous kissing-duplex isomerization. **f,** Native gel showing the reduction in isomerization rate caused by inhibiting exchanging conformations (see Supplementary Fig. 8).

In ES1, the bulge migrates upward by 3 base pairs[43–46]. Here, G29-G28 swap base-pairing partners with A27-G26, A27 swaps with A25, and G26-A25 are bulged out (Fig. 3a). We stabilized ES1 using two point mutants (SL1m(G8C) and SL1m(G8U)) that replace the ES G8•G29 mismatch with the more stable C8•G29 and U8•G29 base pairs (Fig. 3c). Both mutants adopted the ES1 structure as verified by NMR (Supplementary Fig. 4), and relative to wild-type SL1m resulted in large changes in carbon chemical shifts for residues within (G28 and G29) and above (A25, G26 and A27) the internal loop that are directed towards the ES chemical shifts (Fig. 3d and Supplementary Fig. 5). In ES2, the bulge migrates downward by 2 base pairs. Here, G28-G29 swap base-pairing partners with C30-G31, which are now bulged out (Fig. 3a). We stabilized ES2 by replacing the ES2 G7•G28/G8•A27 mismatches with C7•G28/U8•A27 WC base pairs (Fig. 3c). This double mutant (SL1m(G7C/G8U)) adopted the proposed ES2 structure as verified by NMR (Supplementary Fig. 4) and resulted in large changes in the carbon chemical shifts for residues within (G28 and G29) and below (C30 and G31) the internal loop that are directed towards the ES chemical shifts (Fig. 3d). Mutant mimics of ES1 and ES2 induce similar chemical shift perturbations for G28 (C8 and C1′) and G29 (C8), as expected given that they form base pairs in the two cases (Fig. 3d, Supplementary Fig. 5 and Supplementary Discussion). Notably, mutants that stabilize residues above the bulge in their ES conformation also stabilize residues below the bulge in their GS conformation and vice versa (Fig. 3d and Supplementary Fig. 5). This supports the mutual exclusivity of ES1 and ES2 (Fig. 3a); 'trapping' the bulge in the upper (or lower) helix prevents downward (or upward) migration and therefore traps residues in the lower (or upper) helix in their GS.

Together, the GS, ES1 and ES2 define a moving zipper in which bulge residues invade base pairs in the upper or lower helix. Remarkably, an analogous process, if carried out in an intermolecular manner between two SL1 monomers, naturally leads to isomerization and duplex formation most probably through a previously proposed quadruplex-like intermediate[39] (Fig. 3e). Here, bulged-out G28 and G29 can invade base pairs in the upper or lower helix in another monomer to generate ES1- or ES2-like intermolecular base pairs (Fig. 3e). The bulged-out G26 and A25 or C30 and G31 can, in turn, carry out further intermolecular strand invasions, and this process can be repeated to generate a duplex dimer (Fig. 3e). In support of this important role for ES1 and ES2 in SL1 isomerization, mutations that trap ES1 or inhibit formation of ES2 significantly diminish the rate of isomerization, whereas control sequences that preserve the stability of the stem-loop without disrupting conformational exchange show little to no effect (Fig. 3f and Supplementary Fig. 8). Thus, transitions between the GS and ES can promote ATP-independent changes in RNA secondary structure without disrupting the structural integrity of entire hairpins, which may be required for other functions, such as the formation of kissing dimers in SL1.

Compared to secondary structural transitions observed in many regulatory RNA switches[47,48], transitions between the ground and excited states uncovered here involve much more localized changes in RNA structure, occur at rates that are two-to-four orders of magnitude faster, and do not require assistance from external factors. Thus, they can meet unique demands in biological circuits and macromolecular machines. The ESs also present new drug targets and offer new opportunities in the engineering of RNA-based devices. Line-broadening indicative of ESs is routinely observed in NMR spectra of RNA and we therefore predict that RNA ESs exist in great abundance throughout the transcriptome. By combining NMR data with structure prediction tools, it should be possible to determine the three-dimensional structures of RNA ESs at atomic resolution.

METHODS SUMMARY

Detailed methods on RNA sample preparation and assignment, NMR relaxation dispersion data collection and analysis, and isomerization assays can be found in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 13 June; accepted 10 August 2012.

Published online 7 October 2012.

1. Palmer, A. G. & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106,** 1700–1719 (2006).
2. Baldwin, A. J., Hansen, D. F., Vallurupalli, P. & Kay, L. E. Measurement of methyl axis orientations in invisible, excited states of proteins by relaxation dispersion NMR spectroscopy. *J. Am. Chem. Soc.* **131,** 11939–11948 (2009).
3. Neudecker, P. *et al.* Structure of an intermediate state in protein folding and aggregation. *Science* **336,** 362–366 (2012).
4. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).
5. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447,** 1021–1025 (2007).
6. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329,** 1312–1316 (2010).
7. Li, P., Martins, I. R. S., Amarasinghe, G. K. & Rosen, M. K. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nature Struct. Biol.* **15,** 613–618 (2008).
8. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313,** 1638–1642 (2006).
9. Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A. & Al-Hashimi, H. M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J. Am. Chem. Soc.* **131,** 3818–3819 (2009).
10. Massi, F., Johnson, E., Wang, C., Rance, M. & Palmer, A. G. NMR $R_{1\rho}$ rotating-frame relaxation with weak radio frequency fields. *J. Am. Chem. Soc.* **126,** 2247–2256 (2004).
11. Korzhnev, D. M., Orekhov, V. Y. & Kay, L. E. Off-resonance $R_{1\rho}$ NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a Fyn SH3 domain. *J. Am. Chem. Soc.* **127,** 713–721 (2005).
12. Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470,** 498–502 (2011).
13. Hoogstraten, C. G., Wank, J. R. & Pardi, A. Active site dynamics in the lead-dependent ribozyme. *Biochemistry* **39,** 9951–9958 (2000).
14. Johnson, J. E. & Hoogstraten, C. G. Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using C-13 NMR spin relaxation and specific isotope labeling. *J. Am. Chem. Soc.* **130,** 16757–16769 (2008).
15. Blad, H., Reiter, N. J., Abildgaard, F., Markley, J. L. & Butcher, S. E. Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *J. Mol. Biol.* **353,** 540–555 (2005).
16. Dethoff, E. A. *et al.* Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys. J.* **95,** 3906–3915 (2008).
17. Bannwarth, S. & Gatignol, A. HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. *Curr. HIV Res.* **3,** 61–71 (2005).
18. Jaeger, J. A. & Tinoco, I. Jr. An NMR study of the HIV-1 TAR element hairpin. *Biochemistry* **32,** 12522–12530 (1993).
19. Kulinski, T. *et al.* The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. *J. Biol. Chem.* **278,** 38892–38901 (2003).
20. Farès, C., Amata, I. & Carlomagno, T. ^{13}C-detection in RNA bases: revealing structure-chemical shift relationships. *J. Am. Chem. Soc.* **129,** 15814–15823 (2007).
21. Ghose, R., Marino, J. P., Wiberg, K. B. & Prestegard, J. H. Dependence of 13C chemical shifts on glycosidic torsional angles in ribonucleic acids. *J. Am. Chem. Soc.* **116,** 8827–8828 (1994).
22. Nozinovic, S., Furtig, B., Jonker, H. R., Richter, C. & Schwalbe, H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* **38,** 683–694 (2010).
23. Snoussi, K. & Leroy, J.-L. Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry* **40,** 8898–8904 (2001).
24. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452,** 51–55 (2008).
25. Legault, P. & Pardi, A. Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *J. Am. Chem. Soc.* **119,** 6621–6628 (1997).
26. Feng, S. & Holland, E. C. HIV-1 tat trans-activation requires the loop sequence within tar. *Nature* **334,** 165–167 (1988).
27. Berkhout, B. & Jeang, K. T. trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis. *J. Virol.* **63,** 5501–5504 (1989).
28. Richter, S., Cao, H. & Rana, T. M. Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. *Biochemistry* **41,** 6391–6397 (2002).
29. Yoshizawa, S., Fourmy, D. & Puglisi, J. Recognition of the codon-anticodon helix by ribosomal RNA. *Science* **285,** 1722–1725 (1999).
30. Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461,** 1234–1242 (2009).
31. Shandrick, S. *et al.* Monitoring molecular recognition of the ribosomal decoding site. *Angew. Chem. Int. Ed.* **43,** 3177–3182 (2004).
32. Fourmy, D., Recht, M., Blanchard, S. & Puglisi, J. Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274,** 1367–1371 (1996).
33. Romanowska, J., Setny, P. & Trylska, J. Molecular dynamics study of the ribosomal A-site. *J. Phys. Chem. B* **112,** 15227–15243 (2008).
34. O'Connor, M., Thomas, C. L., Zimmermann, R. A. & Dahlberg, A. E. Decoding fidelity at the ribosomal A and P sites: influence of mutations in three different regions of the decoding domain in 16S rRNA. *Nucleic Acids Res.* **25,** 1185–1193 (1997).
35. Dahlquist, K. D. & Puglisi, J. D. Interaction of translation initiation factor IF1 with the *E. coli* ribosomal A site. *J. Mol. Biol.* **299,** 1–15 (2000).
36. Kipper, K., Hetényi, C., Sild, S., Remme, J. & Liiv, A. Ribosomal intersubunit bridge B2a is involved in factor-dependent translation initiation and translational processivity. *J. Mol. Biol.* **385,** 405–422 (2009).
37. Moore, M. D. & Hu, W.-S. HIV-1 RNA dimerization: It takes two to tango. *AIDS Rev.* **11,** 91–102 (2009).
38. Clever, J. L. & Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71,** 3407–3414 (1997).
39. Rist, M. J. & Marino, J. P. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **41,** 14762–14770 (2002).
40. Mujeeb, A. *et al.* Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acids Res.* **35,** 2026–2034 (2007).
41. Turner, K. B., Hagan, N. A. & Fabris, D. Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein. *J. Mol. Biol.* **369,** 812–828 (2007).
42. Takahashi, K. *et al.* Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **6,** 96–102 (2000).
43. Sun, X., Zhang, Q. & Al-Hashimi, H. M. Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg^{2+} binding: role in the kissing-duplex structural transition. *Nucleic Acids Res.* **35,** 1698–1713 (2007).
44. Yuan, Y., Kerwood, D. J., Paoletti, A. C., Shubsda, M. F. & Borer, P. N. Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry* **42,** 5259–5269 (2003).
45. Lawrence, D. C., Stover, C. C., Noznitsky, J., Wu, Z. & Summers, M. F. Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem-loop SL1. *J. Mol. Biol.* **326,** 529–542 (2003).
46. Ulyanov, N. B. NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J. Biol. Chem.* **281,** 16168–16177 (2006).
47. Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43,** 867–879 (2011).
48. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* **482,** 322–330 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements E.A.D., K.P. and J.C contributed equally to this work. We thank members of the Al-Hashimi laboratory for input. We acknowledge the Michigan Economic Development Cooperation and the Michigan Technology Tri-Corridor for the support of the purchase of a 600 MHz spectrometer. K.P. is supported by a postdoctoral Fellowship from the Swedish Research Council (VR-K2011-78PK-21662-0-12). This work was supported by the US National Institutes of Health (R01 AI066975) and by a Rackham Graduate Student Research Grant awarded by the University of Michigan.

Author Contributions H.M.A., E.A.D., K.P. and J.C. conceived the approaches to structurally characterize RNA ES and wrote the paper. E.A.D. and K.P. performed all experiments and data analyses for HIV TAR and SL1m, respectively. J.C. with assistance from A.C.-N. performed all experiments and data analyses for the A-site.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.M.A. (hashimi@umich.edu).

doi:10.1038/nature11592

An over-massive black hole in the compact lenticular galaxy NGC 1277

Remco C. E. van den Bosch[1,2], Karl Gebhardt[2], Kayhan Gültekin[3], Glenn van de Ven[1], Arjen van der Wel[1] & Jonelle L. Walsh[2]

Most massive galaxies have supermassive black holes at their centres, and the masses of the black holes are believed to correlate with properties of the host-galaxy bulge component[1]. Several explanations have been proposed for the existence of these locally established empirical relationships, including the non-causal, statistical process of galaxy–galaxy merging[2], direct feedback between the black hole and its host galaxy[3], and galaxy–galaxy merging and the subsequent violent relaxation and dissipation[4]. The empirical scaling relations are therefore important for distinguishing between various theoretical models of galaxy evolution[5,6], and they furthermore form the basis for all black-hole mass measurements at large distances. Observations have shown that the mass of the black hole is typically 0.1 per cent of the mass of the stellar bulge of the galaxy[7,8]. Until now, the galaxy with the largest known fraction of its mass in its central black hole (11 per cent) was the small galaxy NGC 4486B[1,9]. Here we report observations of the stellar kinematics of NGC 1277, which is a compact, lenticular galaxy with a mass of 1.2×10^{11} solar masses. From the data, we determine that the mass of the central black hole is 1.7×10^{10} solar masses, or 59 per cent of its bulge mass. We also show observations of five other compact galaxies that have properties similar to NGC 1277 and therefore may also contain over-massive black holes. It is not yet known if these galaxies represent a tail of a distribution, or if disk-dominated galaxies fail to follow the usual black-hole mass scaling relations[4,10].

Direct measurements of black-hole mass often rely on obtaining spatially resolved stellar or gas kinematics within the black hole's 'sphere of influence', that is, the region over which it dominates the gravitational potential. We have obtained long-slit spectroscopy of 700 nearby galaxies with the Marcario Low Resolution Spectrograph[11] on the Hobby-Eberly Telescope, Texas, to find suitable targets for direct measurements of black-hole mass (Supplementary Information). As shown in Table 1, six of these galaxies have very peculiar properties; they have velocity dispersions of $\sigma > 350\,\mathrm{km\,s^{-1}}$ and half-light radii of $R_e < 3\,\mathrm{kpc}$. It is unusual for such small galaxies to have such large dispersions, which signify unusually high central mass concentrations: a simple virial mass estimate indicates that the central 200 pc of the galaxies listed in Table 1 contains more than 10 billion solar masses, which is 100 times more than typical galaxies of the same size.

Black-hole masses can be measured directly by fitting self-consistent Schwarzschild models[12] to spatially resolved spectroscopy data and high-resolution imaging. Archival Hubble Space Telescope (HST) imaging is available for one of these six dense galaxies, NGC 1277. On the basis of the HST imaging (Fig. 1) and the stellar kinematics (Fig. 2), we constructed 600,000 orbit-based models using iterative refinement to search parameter space[13,14]. The best-fit model is then found by marginalizing over all parameters: the stellar mass-to-light ratio, the black-hole mass and the mass and concentration of the Navarro–Frenk–White dark halo[15]. The confidence intervals are determined with the goodness-of-fit statistic χ^2. We measure a black-hole mass of $(17 \pm 3) \times 10^9$ solar masses ($M_\odot$) and a total stellar mass of $(1.2 \pm 0.4) \times 10^{11} M_\odot$, with 1-s.d. confidence intervals based on $\Delta\chi^2 = 1$ after marginalizing over the dark-halo parameters (Supplementary Information). The black hole in NGC 1277 is one of the most massive black holes to be dynamically confirmed, and moreover has a mass fraction of 14% of the total stellar mass in the galaxy.

No galaxy with such a large ratio of black-hole mass to stellar mass has previously been seen. Owing to the strong disk-like rotation (Fig. 2) and the lack of an unambiguous bulge in NGC 1277 (Fig. 1), it is unclear where to evaluate its black hole against the relation between black-hole mass and bulge luminosity. The central pseudobulge contains 24% of the light (Fig. 1) and the black-hole/bulge mass fraction is 59%. As shown in Fig. 3, NGC 1277 is a significant outlier from the mass–luminosity relation, by two orders of magnitude. At a fixed bulge luminosity of $3 \times 10^{10} L_{K,\odot}$ to $10 \times 10^{10} L_{K,\odot}$, where $L_{K,\odot}$ is the K-band solar luminosity, dynamical measurements of black-hole mass now range over four orders of magnitude, from $10^6 M_\odot$ to $10^{10} M_\odot$, showing that bulge (or pseudobulge) luminosity is not a good predictor of black-hole mass.

We now place NGC 1277 on the relation between black-hole mass ($M_\bullet$) and velocity dispersion. The average velocity dispersion inside the half-light radius (2.8″) and outside the sphere of influence (1.6″) for NGC 1277 is $\sigma = 333\,\mathrm{km\,s^{-1}}$, according to a reconstruction of the best-fit orbit-based model. For this value of σ, the most recent inferred

Table 1 | Global properties of the six compact, high-dispersion galaxies

Object	Distance (Mpc)	σ_c (km s^{-1})	$R_{e,K}$ (kpc)	$\log(L_{K,\odot})$	$\varepsilon\,(1-b/a)$
ARK 90	131	392 ± 4	1.6	11.2	0.7
NGC 1270	69	393 ± 3	1.8	11.2	0.8
NGC 1277	73	403 ± 4	1.6	11.1	0.5
UGC 1859	82	362 ± 4	2.0	11.2	0.6
UGC 2698	89	397 ± 3	2.7	11.4	0.7
MRK 1216	94	354 ± 4	1.9	11.2	0.6

The six galaxies presented here were observed with the Marcario Low Resolution Spectrograph[11] as part of a large survey programme. We targeted galaxies from the Two Micron All Sky Survey (2MASS) extended-source catalogue[24] that are expected to have the largest spheres of influence. Our predictions of sphere of influence assume that the galaxies follow the relation between black-hole mass and host-galaxy velocity dispersion[1]. For those 2MASS galaxies without a known velocity dispersion value, we used an estimate based on the fundamental-plane relation between galaxy size, surface brightness and velocity dispersion[25]. See the Supplementary Information for more information on the survey. The columns show the following near-infrared properties: distance from Hubble flow (column 2); stellar velocity dispersion extracted from a central aperture (column 3); and 2MASS[24] half-light radius (K band; column 4), total luminosity (K band; column 5) and apparent ellipticity (column 6), where *a* and *b* are respectively the lengths of the apparent major and minor axes.

---

[1]Max-Planck Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany. [2]Department of Astronomy, The University of Texas at Austin, 1 University Station C1400, Austin, Texas 78712, USA. [3]Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA.

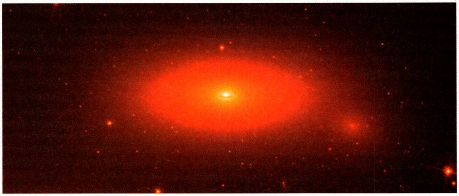


Figure 1 | Optical HST image of the compact lenticular galaxy NGC 1277. The image is scaled with the logarithm of the luminosity and is 19 kpc × 8 kpc, with north pointing up and east to the left. In this high-resolution optical image, the galaxy has a half-light radius of 1 kpc, is strongly flattened and is disky. It is clear that a superposition of multiple galaxies does not explain the high velocity dispersion. NGC 1277 has a small, regular, nuclear dust disk with an apparent axis ratio of only 0.3, which indicates that we see the galaxy close to edge-on. Through a multicomponent fit[26] to the HST image, we identify the inner component, with a half-light radius of 0.3 kpc and a Sérsic index of $n \approx 1$, as a pseudobulge that contains 24% of the light. For the dynamical modelling, we construct a three-dimensional luminous-mass model of the stars by de-projecting the two-dimensional light model from the HST image. Then the gravitational potential is inferred from the combined luminous, black-hole and dark-matter halo mass distribution. In this potential, representative orbits are integrated numerically, keeping track of the path and velocity along each orbit. We then create a reconstruction of the galaxy by assigning each orbit an amount of light, such that the model recreates the total light distribution, while simultaneously fitting the long-slit stellar kinematics observed with the Hobby-Eberly Telescope (Fig. 2). The models include the effect of the Earth's atmosphere and the telescope optics without any a-priori assumption on the orbital configuration (Supplementary Information).

relation[16] predicts a black-hole mass of $2.4 \times 10^9 M_\odot$, so the measured value is almost one order of magnitude higher, or a 2.1-s.d. outlier relative to the intrinsic scatter in the $M_\bullet$–σ relation[1].

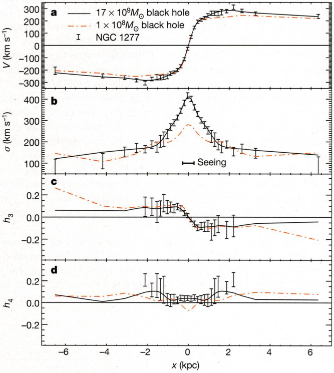


Figure 2 | Line-of-sight stellar kinematics of NGC 1277. The stellar kinematics as observed with the Marcario Low Resolution Spectrograph[11], shown with 1-s.d. error bars, and measured[27] at 31 locations along the major axis of NGC 1277 (Supplementary Information): mean velocity (**a**); velocity dispersion (**b**); and higher-order Gauss–Hermite velocity moments[28] h_3 (**c**) and h_4 (**d**), respectively representing skewness and kurtosis. The kinematics show remarkably strong rotation and a dispersion profile that strongly peaks towards the centre. The best-fit Schwarzschild model (black line) corresponds to a $17 \times 10^9 M_\odot$ black hole. The relation between black-hole mass and host luminosity predicts a $10^8 M_\odot$ black hole, but the corresponding model (red dot–dash line) does not fit the data at all. The telescope resolution (seeing full-width at half-maximum, 1.6″) is indicated in **b** and is sufficient to resolve the sphere of influence of the black hole.

Apart from NGC 1277, NGC 4486B[9] and Henize 2-10[17] are known to lie significantly above the relations, and at least three galaxies are known to lie significantly below the relations[10,18,19]. We do not yet know if these over-massive and under-massive black holes just lie in the tails of a relatively narrow distribution of joint black-hole/galaxy properties, or if they demonstrate non-universality. Only through more black-hole measurements, including those in the other five compact galaxies with high velocity dispersions, we will be able to establish the cause of the black-hole/galaxy connection.

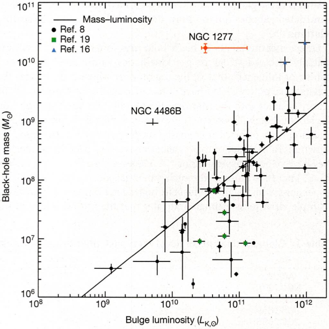


Figure 3 | The correlation between black-hole mass and near-infrared bulge luminosity, $L_{K,\odot}$. The black line shows the mass–luminosity relation[8] for galaxies with a directly measured black-hole mass. NGC 1277 is a significant positive outlier. In addition to the galaxies (black dots) to which the relation has been fitted[8], eight black-hole masses (NGC 4486B[9], triangles[16], squares[19]) have been added with 2MASS K-band bulge luminosities. The error bars denote 1-s.d. uncertainties, except for the NGC 1277 bulge luminosity, where we use its total luminosity as a conservative upper limit.

A stellar population analysis of NGC 1277[20] indicates that it contains only old ($\gtrsim 8$ Gyr) stars and that there has not been any recent star formation. The black hole must thus have been in place for at least 8 Gyr, because black-hole accretion without much star formation or the formation of a (classical) bulge is highly unlikely. Furthermore, there is no strong evidence that NGC 1277 has been tidally stripped, because its isophotes are extremely regular and disky, it seems to have a normal dark-matter halo as inferred from the dynamical model, and at large radii the rotation curve is flat out to five times the half-light radius.

Although the six compact galaxies presented in Table 1 are unusual in the present-day universe, they are quantitatively similar to the typical red, passive galaxies at much earlier times (at redshifts of $z \approx 2$): those are also found, on average, to be smaller than similarly massive galaxies in the present-day universe[21], possibly possess high velocity dispersions[22] and generally have a disk-like structure[23]. Perhaps the compact systems we found are local analogues of these high-redshift galaxies.

Received 4 May; accepted 13 September 2012.

1. Gültekin, K. *et al.* The M-σ and M-L relations in galactic bulges, and determinations of their intrinsic scatter. *Astrophys. J.* **698,** 198 (2009).
2. Jahnke, K. & Macciò, A. V. The non-causal origin of the black-hole-galaxy scaling relations. *Astrophys. J.* **734,** 92 (2011).
3. Fabian, A. C. The obscured growth of massive black holes. *Mon. Not. R. Astron. Soc.* **308,** L39–L43 (1999).
4. Kormendy, J., Bender, R. & Cornell, M. E. Supermassive black holes do not correlate with galaxy disks or pseudobulges. *Nature* **469,** 374–376 (2011).
5. Croton, D. J. *et al.* The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365,** 11–28 (2006).
6. Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E. & Hernquist, L. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *Mon. Not. R. Astron. Soc.* **391,** 481–506 (2008).
7. Häring, N. & Rix, H.-W. On the black hole mass-bulge mass relation. *Astrophys. J.* **604,** L89–L92 (2004).
8. Sani, E., Marconi, A., Hunt, L. K. & Risaliti, G. The Spitzer/IRAC view of black hole-bulge scaling relations. *Mon. Not. R. Astron. Soc.* **413,** 1479–1494 (2011).
9. Magorrian, J. *et al.* The demography of massive dark objects in galaxy centers. *Astron. J.* **115,** 2285–2305 (1998).
10. Nowak, N. *et al.* Do black hole masses scale with classical bulge luminosities only? The case of the two composite pseudo-bulge galaxies NGC 3368 and NGC 3489. *Mon. Not. R. Astron. Soc.* **403,** 646–672 (2010).
11. Hill, G. J. *et al.* Hobby-Eberly Telescope low-resolution spectrograph. *Proc. SPIE* **3355,** 375–386 (1998).
12. Schwarzschild, M. A numerical model for a triaxial stellar system in dynamical equilibrium. *Astrophys. J.* **232,** 236–247 (1979).
13. van den Bosch, R. C. E., van de Ven, G., Verolme, E. K., Cappellari, M. & de Zeeuw, P. T. Triaxial orbit based galaxy models with an application to the (apparent) decoupled core galaxy NGC 4365. *Mon. Not. R. Astron. Soc.* **385,** 647–666 (2008).
14. van den Bosch, R. C. E. & de Zeeuw, P. T. Estimating black hole masses in triaxial galaxies. *Mon. Not. R. Astron. Soc.* **401,** 1770–1780 (2010).
15. Navarro, J. F., Frenk, C. S. & White, S. D. M. The structure of cold dark matter halos. *Astrophys. J.* **462,** 563 (1996).
16. McConnell, N. J. *et al.* Two ten-billion-solar-mass black holes at the centres of giant elliptical galaxies. *Nature* **480,** 215–218 (2011).
17. Reines, A. E., Sivakoff, G. R., Johnson, K. E. & Brogan, C. L. An actively accreting massive black hole in the dwarf starburst galaxy Henize 2-10. *Nature* **470,** 66–68 (2011).
18. Merritt, D., Ferrarese, L. & Joseph, C. L. No supermassive black hole in M33? *Science* **293,** 1116–1118 (2001).
19. Greene, J. E. *et al.* Precise black hole masses from megamaser disks: black hole-bulge relations at low mass. *Astrophys. J.* **721,** 26–45 (2010).
20. Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G. & Gomes, J. M. Semi-empirical analysis of Sloan Digital Sky Survey galaxies – I. Spectral synthesis method. *Mon. Not. R. Astron. Soc.* **358,** 363–378 (2005).
21. van Dokkum, P. G. *et al.* Confirmation of the remarkable compactness of massive quiescent galaxies at $z \sim 2.3$: early-type galaxies did not form in a simple monolithic collapse. *Astrophys. J.* **677,** L5–L8 (2008).
22. van Dokkum, P. G., Kriek, M. & Franx, M. A high stellar velocity dispersion for a compact massive galaxy at redshift $z = 2.186$. *Nature* **460,** 717–719 (2009).
23. van der Wel, A. *et al.* The majority of compact massive galaxies at $z \sim 2$ are disk dominated. *Astrophys. J.* **730,** 38 (2011).
24. Jarrett, T. H. *et al.* 2MASS extended source catalog: overview and algorithms. *Astron. J.* **119,** 2498–2531 (2000).
25. Pahre, M. A., Djorgovski, S. G. & de Carvalho, R. R. Near-infrared imaging of early-type galaxies. III. The near-infrared fundamental plane. *Astron. J.* **116,** 1591–1605 (1998).
26. Peng, C. Y., Ho, L. C., Impey, C. D. & Rix, H.-W. Detailed structural decomposition of galaxy images. *Astron. J.* **124,** 266–293 (2002).
27. Cappellari, M. & Emsellem, E. Parametric recovery of line-of-sight velocity distributions from absorption-line spectra of galaxies via penalized likelihood. *Publ. Astron. Soc. Pacif.* **116,** 138–147 (2004).
28. van der Marel, R. P. & Franx, M. A new method for the identification of non-Gaussian line profiles in elliptical galaxies. *Astrophys. J.* **407,** 525–539 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements K. Gebhardt and J.L.W. are supported by the US National Science Foundation (NSF-0908639, AST-1102845). K. Gültekin acknowledges support provided by the US National Aeronautics Space Administration (GO0-11151X, GO2-13111X) and the Space Telescope Science Institute (HST-GO-12557.01-A). The Hobby-Eberly Telescope is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München and Georg-August-Universität Göttingen. The Hobby-Eberly Telescope is named in honour of its principal benefactors, William P. Hobby and Robert E. Eberly.

Author Contributions R.C.E.v.d.B. designed the survey and carried out the data analysis and the modelling. R.C.E.v.d.B. and G.v.d.V. wrote the manuscript. A.v.d.W. carried out the image analysis. All authors contributed to the interpretation of the observations and the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.C.E.v.d.B. (bosch@mpia.de).

LETTER

doi:10.1038/nature11611

Active upper-atmosphere chemistry and dynamics from polar circulation reversal on Titan

Nicholas A. Teanby[1], Patrick G. J. Irwin[2], Conor A. Nixon[3], Remco de Kok[4], Sandrine Vinatier[5], Athena Coustenis[5], Elliot Sefton-Nash[1,6], Simon B. Calcutt[2] & F. Michael Flasar[3]

Saturn's moon Titan has a nitrogen atmosphere comparable to Earth's, with a surface pressure of 1.4 bar. Numerical models reproduce the tropospheric conditions very well but have trouble explaining the observed middle-atmosphere temperatures, composition and winds[1,2]. The top of the middle-atmosphere circulation has been thought to lie at an altitude of 450 to 500 kilometres, where there is a layer of haze that appears to be separated from the main haze deck[3]. This 'detached' haze was previously explained as being due to the colocation of peak haze production and the limit of dynamical transport by the circulation's upper branch[4]. Here we report a build-up of trace gases over the south pole approximately two years after observing the 2009 post-equinox circulation reversal, from which we conclude that middle-atmosphere circulation must extend to an altitude of at least 600 kilometres. The primary drivers of this circulation are summer-hemisphere heating of haze by absorption of solar radiation and winter-hemisphere cooling due to infrared emission by haze and trace gases[5]; our results therefore imply that these effects are important well into the thermosphere (altitudes higher than 500 kilometres). This requires both active upper-atmosphere chemistry, consistent with the detection of high-complexity molecules and ions at altitudes greater than 950 kilometres[6,7], and an alternative explanation for the detached haze, such as a transition in haze particle growth from monomers to fractal structures[8].

Saturn's 26.7° obliquity means that Titan's atmosphere experiences large solar flux variations during its 29.5-yr orbit around the Sun. For most of Titan's year, middle-atmosphere (stratosphere and mesosphere at altitudes between 100 and 500 km) circulation is predicted to comprise a single pole-to-pole circulation cell, with summer-hemisphere upwelling, winter-hemisphere subsidence and a winter-hemisphere circumpolar vortex[1,2,5,9–12]. This was confirmed during northern winter using measurements of temperature and trace-gas abundance made by NASA's Cassini spacecraft[13–18]. Titan experienced northern spring equinox on 11 August 2009, around which time changes in solar flux distribution were predicted to cause a reversal of the middle-atmospheric circulation. Such dynamical changes can be probed using profiles of temperature and trace-gas abundance[15] derived from infrared spectra measured with Cassini's Composite Infrared Spectrometer[19] (CIRS). Therefore, to investigate the reversal mechanism we analysed all available south polar limb (horizontal viewing) CIRS observations made in the 4-yr period centred on the equinox. This included measurements at complementary high (0.5 cm^{-1}) and low (14 cm^{-1}) spectral resolutions (Supplementary Information, Supplementary Fig. 1 and Supplementary Table 1).

The CIRS observations show that a very large increase in high-altitude trace-gas emission occurred over the south pole sometime between late 2010 and mid 2011 (Fig. 1). From both high- and low-resolution observation sequences, we derived altitude profiles of temperature, HCN, HC_3N and C_2H_2 using a nonlinear optimal estimation inversion method[20] that closely follows our previous studies[14,21] (Supplementary Information and Supplementary Table 2). Additionally, the high-resolution data allowed determination of C_3H_4, C_4H_2 and C_6H_6 profiles whose emission peaks were too close together to be resolved in the low-resolution data.

Inversion results show very rapid changes in atmospheric temperature and composition, especially after the equinox (Figs 2–4 and Supplementary Fig. 2). The observed south polar warming at altitudes above 300 km suggests that subsidence over the south pole is initiated just after equinox, with the increased temperature being due to adiabatic heating as upper atmosphere air is advected to higher pressures and compressed. This is similar to the process that caused a subsidence-induced north polar hotspot during northern winter[13,22]. The observed temperature structure implies subsidence velocities of 0.5–2.0 mm s^{-1}, broadly consistent with predictions from numerical models (Supplementary Information and Supplementary Table 3). Subsidence is weakest just after equinox in early 2010, at 0.5 mm s^{-1}, but quickly increases to 1.5 mm s^{-1} by June 2010 (2010.43). Cooling observed in the stratosphere (150–300 km) suggests that for the period covered by our data subsidence does not extend to lower altitudes. The cooling by 20 K that is evident between January 2010 (2010.04) and September 2011 (2011.70) at altitudes of 150–300 km (Fig. 4) is most likely due to radiative cooling from the lower atmosphere, which since equinox has been experiencing reduced insolation, and is consistent with the $\sim$1-yr cooling timescale at these altitudes[23].

Changes in upper-atmosphere composition occur on similarly short timescales (Figs 3 and 4 and Supplementary Fig. 2), with evidence of large increases in trace-gas abundances occurring in 2011. This can be explained by a combination of subsidence and photochemically induced vertical gradients. High-altitude (>500 km) photochemical reactions produce trace compounds such as HCN and HC_3N, which are transported into the lower atmosphere by vertical mixing processes, where they are destroyed by photolysis or removed by condensation near the tropopause[24]. The result is increasing relative abundances of these compounds with altitude and a vertical gradient inversely proportional to species lifetime[25]. Subsidence would advect these profiles downwards, causing enrichment at lower atmospheric levels[15] and explaining the observed increase.

Therefore, both observed temperature and abundance increases are consistent with mesospheric (>300 km) south polar subsidence during the post-equinox period. This implies a recent reversal in circulation direction for the south polar mesosphere, relative to the circulation direction derived earlier in the mission[13–16]. An alternative, purely radiative, explanation for the temperature changes can be rejected. Radiative time constants at mesospheric altitudes are short relative to Titan's seasons, and temperature should thus in theory be able to react rapidly to changes in seasonal solar flux. However, photochemical lifetimes of most trace gases are comparable to or greater

---

[1] School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK. [2] Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK. [3] Planetary Systems Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [4] SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands. [5] LESIA Observatoire de Paris, CNRS, UPMC Université Paris 06, Université Paris-Diderot, 5 place Jules Janssen, 92195 Meudon Cedex, France. [6] Department of Earth and Space Sciences, University of California Los Angeles, Los Angeles, California 90095-1567, USA.

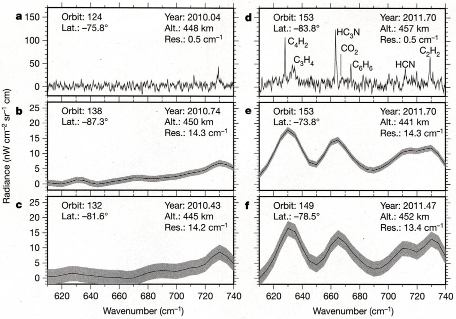


Figure 1 | Rapid south polar atmospheric change observed using infrared spectra. These observations were made using Cassini CIRS after the 11 August 2009 equinox and are grouped as follows: before 1 January 2011 (**a–c**); after 1 January 2011 (**d–f**). The spectra indicate that between late 2010 and early 2011 there was a large increase in trace-gas emission at the south pole. This is observed in three independent observation sequences at both high and low spectral resolution. Altitude refers to the tangent altitude, which is the closest distance to Titan's surface reached by the line-of-sight vector, and is approximately 450 km for these spectra. Grey areas indicate measurement error envelopes (s.e.). We focused on the 610–740 cm^{-1} (16.4–13.5 μm) spectral region, which contains strong trace-gas emission features. Since the 2009 northern spring equinox, Cassini remained in an equatorial orbit around Saturn, which was ideal for limb sounding (horizontal viewing), and many limb measurements of the south polar region were taken (Supplementary Table 1). Most observations were of a single latitude, but several limb-mapping sequences were also measured, covering multiple latitudes at a time and allowing the determination of latitude–altitude cross-sections through the atmosphere (Figs 2 and 3).

than seasonal timescales. Therefore, to be consistent with all our data, the observed changes must be due to a reversal of the circulation as opposed to changes in direct solar heating.

We note that, whereas the January 2010 (2010.04) temperature results show polar warming at high altitudes almost immediately after equinox, there is no evidence for large increases in trace-gas abundances until much later. In fact, the first evidence for increases in south polar trace gas is in June 2011 (2011.47) (Fig. 3), and this is corroborated by subsequent observations in September 2011 (2011.70) (Fig. 4 and Supplementary Fig. 2). However, unlike increases in south polar temperature caused by adiabatic heating, trace gases take time to advect from upper-atmosphere source regions to observable altitudes, which means that the temporal offset between temperature and composition results is not inconsistent. Our results suggest that this advection process takes approximately 1.5–2 yr after reversal initiation. This corresponds to ~100 km of polar subsidence, assuming the 1.5 mm s^{-1} subsidence rate inferred from polar temperature anomalies.

An independent check on this interpretation and on south polar subsidence rates can be obtained from the composition results themselves. Polar abundance increases at 450 km are at least an order of magnitude for all gases (Fig. 4 and Supplementary Table 4) except C_2H_2, which has a more muted response in keeping with its longer atmospheric lifetime and shallower vertical gradient. A first-order approximation, combining results from all gases and assuming no photochemical alteration of gas profiles, implies average subsidence velocities between January 2010 (2010.04) and September 2011

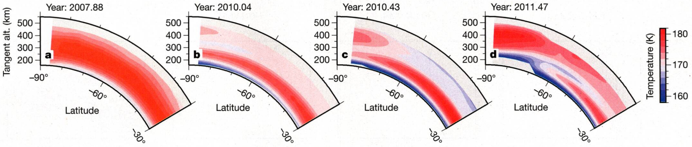


Figure 2 | South polar seasonal temperature changes. Cross-sections were derived from low-spectral-resolution limb-mapping sequences and cover pre-equinox (**a**) and post-equinox (**b–d**) periods. Substantial stratospheric (<300 km) cooling occurs after the equinox, consistent with reduced total solar flux during this time[17,23], as Titan moves towards southern winter. After the equinox (after mid-2009), there is evidence for high-altitude (450 km) polar warming relative to more equatorial latitudes. This is initially present as a small (2 K) temperature anomaly almost immediately proceeding the equinox (**b**), which increases to 6 K (**c**) and then to 8 K (**d**) in subsequent sequences. This implies that the mesospheric circulation has reversed and is now subsiding at the south pole. The strongest polar warming occurs in the most recent observation, indicating the fastest subsidence speeds. Grey regions indicate latitudes and altitudes where observations exist but have insufficient signal-to-noise ratios for an accurate temperature determination. Contour spacing is 2 K, which is the maximum uncertainty for this altitude range. These changes are confirmed by additional single-latitude observations at both high (Fig. 4) and low (Supplementary Fig. 2) spectral resolution.

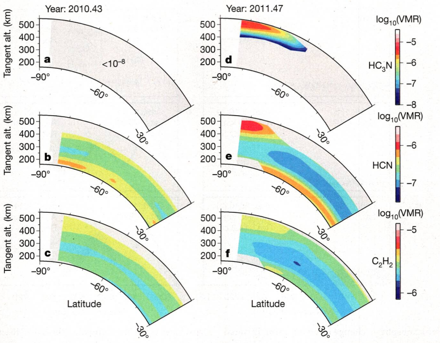


Figure 3 | Seasonal changes in south polar trace-gas abundances. Cross-sections were derived from low-spectral-resolution limb-mapping sequences and show results for June 2010 (2010.43) (**a**–**c**) and June 2011 (2011.47) (**d**–**f**). In the 2011 observation, trace-gas abundances have increased substantially at high altitudes (>450 km) over the south pole. The most pronounced increases occur for HCN and HC_3N. This is consistent with temperature determinations in Fig. 2 and implies a reversal in mesospheric circulation, with subsidence now occurring at the south pole. Grey regions indicate latitudes and altitudes where observations exist but have insufficient signal-to-noise ratios for an accurate abundance determination. We note that HC_3N cannot be reliably determined in June 2010 (2010.43) owing to its very low relative abundance ($<10^{-8}$). Low-resolution mapping sequences taken before June 2010 (2010.43) show comparable compositions to those in **a**–**c**. VMR is the volume mixing ratio and quantifies the relative atmospheric abundance of each species.

(2011.70) of 0.8–2.3 mm s^{-1} (Supplementary Information and Supplementary Table 4). This is in excellent agreement with values of 0.5–2.0 mm s^{-1} derived from the temperature results.

The mechanism for reversal of middle-atmosphere circulation is related to solar flux distribution and angular momentum transfer. Stratospheric temperatures are not symmetric at northern spring equinox, but are slightly warmer in the south[17,23,26]. Therefore, given that temperatures and zonal winds are coupled by the thermal wind equation, during springtime the atmosphere has to transport angular momentum from the pole leaving winter, where the circumpolar

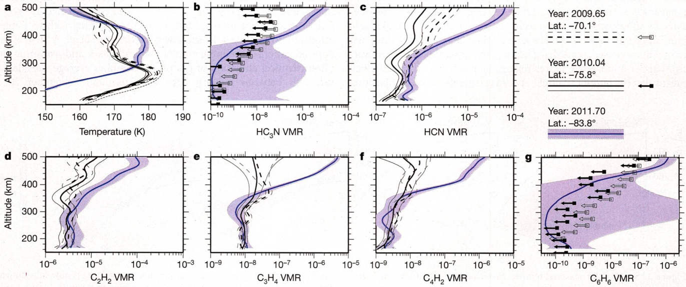


Figure 4 | Profiles of temperature and composition taken close to the equinox. All profiles were derived from high-spectral-resolution (0.5 cm^{-1}), single-latitude limb integrations. **a**, Temperature profiles show evidence for mesospheric polar warming, which increases and moves to lower altitude with time. **b**–**g**, Large increases in all trace-gas species are visible in September 2011 (2011.70). Thin dashed lines, thin solid lines, and light blue shading indicate the error envelope (s.e.) for August 2009 (2009.65), January 2010 (2010.04) and September 2011 (2011.70) observations, respectively. The short-dashed line in **a** indicates the a-priori profile used to start the inversion. HC_3N (**b**) and C_6H_6 (**g**) are not reliably detected in August 2009 (2009.65) or January 2010 (2010.04), so 1σ upper limits are given instead (arrows). These high-spectral-resolution observations confirm the trends seen in the low-spectral-resolution mapping sequences (Figs 2 and 3).

winds are strong, to the pole moving towards winter, where the winds are weak[26]. This results in a cross-equatorial circulation from north to south at high altitudes, driving subsidence in the south polar atmosphere and explaining the observed adiabatic heating and increased trace-gas abundances. Our observations of temperature now constrain the mesospheric reversal timing to shortly after equinox—perhaps coincident with it, but certainly within six months (0.015 Titan years). Furthermore, at present the observed spatial distribution of abundance and temperature increases limits the main zone of subsidence to 90–70° S, with subsidence velocities of 0.5–2.3 mm s^{-1}.

The observed reversal timing agrees very well with recent atmospheric general circulation models[1,2,11]. Unfortunately, direct comparison of observed temperature and composition with results of these models is not possible at the moment because, as noted previously[1,2], the models have a relatively low upper boundary, which means that whereas our peak seasonal signal occurs above 400 km, the models are only applicable at lower altitudes. Decoupling of tropospheric and stratospheric circulation means that this has limited effect on the lowermost atmosphere, and there is generally good agreement between models and observations for the non-superrotating troposphere and the surface[2]. However, the low model top distorts the middle-atmosphere structure and places the polar zonal jets at too low an altitude compared with observational constraints[1,2,23]. Our results show that the reversal in atmospheric circulation takes place throughout the mesosphere, with temperatures implying subsidence in the 300–500-km altitude range and composition implying subsidence up to an altitude of at least 600 km, into photochemical source regions, which is necessary to provide the high trace-gas abundance we observe. Therefore, although the agreement between numerical models and observations is encouraging in terms of reversal timing and approximate reproduction of key atmospheric features such as superrotation and zonal jets, our results show that it is critical for the next generation of models to be extended to higher altitudes to fully capture Titan's dynamical behaviour.

If the circulation does extend to 600 km altitude or more then the detached haze layer observed around 450–500 km cannot mark the top of middle-atmosphere circulation, as has been previously assumed. Recent seasonal changes in the altitude of the detached haze[27] provide a strong argument for circulation-induced modification of the haze, but our observations show that an origin for the detached haze in terms of dynamical transport and a coincident peak haze-production altitude[4] cannot provide a complete explanation. Instead, an origin in a transition from monomer to fractal haze particles[8], combined with higher-altitude haze production and subsequent modification by dynamical circulation, is required. Such high-altitude circulation also needs a driving mechanism, implying that solar heating of haze and cooling due to infrared emission from trace gases is important at higher altitudes than previously thought.

Therefore, a consistent picture of Titan's middle and upper atmospheres is now emerging. Complex chemistry occurs in the uppermost atmosphere, as evidenced by heavy ions and molecules detected by Cassini's *in situ* instruments[6,7,28] and ultraviolet observations of haze opacity in the thermosphere[29], in broad agreement with active thermospheric photochemistry predicted by the most recent one-dimensional photochemical models[24,30]. Our measurements show that the radiative effects of this complex chemistry are sufficient to drive dynamics up to very high altitudes, effectively linking chemical and dynamical processes well into the thermosphere (>500 km).

Received 30 March; accepted 19 September 2012.

1. Newman, C. E., Lee, C., Lian, Y., Richardson, M. I. & Toigo, A. D. Stratospheric superrotation in the TitanWRF model. *Icarus* **213,** 636–654 (2011).
2. Lebonnois, S., Burgalat, J., Rannou, P. & Charnay, B. Titan global climate model: a new 3-dimensional version of the IPSL Titan GCM. *Icarus* **218,** 707–722 (2012).
3. Porco, C. C. *et al.* Imaging of Titan from the Cassini spacecraft. *Nature* **434,** 159–168 (2005).
4. Rannou, P., Hourdin, F. & McKay, C. P. A wind origin for Titan's haze structure. *Nature* **418,** 853–856 (2002).
5. Hourdin, F. *et al.* Numerical simulation of the general circulation of the atmosphere of Titan. *Icarus* **117,** 358–374 (1995).
6. Waite, J. H. *et al.* Ion neutral mass spectrometer results from the first flyby of Titan. *Science* **308,** 982–986 (2005).
7. Coates, A. J. *et al.* Discovery of heavy negative ions in Titan's ionosphere. *Geophys. Res. Lett.* **34,** L22103 (2007).
8. Lavvas, P. P., Yelle, R. V. & Vuitton, V. The detached haze layer in Titan's mesosphere. *Icarus* **201,** 626–633 (2009).
9. Lebonnois, S., Toublanc, D., Hourdin, F. & Rannou, P. Seasonal variations of Titan's atmospheric composition. *Icarus* **152,** 384–406 (2001).
10. Hourdin, F., Lebonnois, S., Luz, D. & Rannou, P. Titan's stratospheric composition driven by condensation and dynamics. *J. Geophys. Res.* **109,** E12005 (2004).
11. Rannou, P., Lebonnois, S., Hourdin, F. & Luz, D. Titan atmosphere database. *Adv. Space Res.* **36,** 2194–2198 (2005).
12. Crespin, A. *et al.* Diagnostics of Titan's stratospheric dynamics using Cassini/CIRS data and the 2-dimensional IPSL circulation model. *Icarus* **197,** 556–571 (2008).
13. Flasar, F. M. *et al.* Titan's atmospheric temperatures, winds, and composition. *Science* **308,** 975–978 (2005).
14. Teanby, N. A. *et al.* Titan's winter polar vortex structure revealed by chemical tracers. *J. Geophys. Res.* **113,** E12003 (2008).
15. Teanby, N. A., Irwin, P. G. J., de Kok, R. & Nixon, C. A. Dynamical implications of seasonal and spatial variations in Titan's stratospheric composition. *Phil. Trans. R. Soc. Lond. A* **367,** 697–711 (2009).
16. Coustenis, A. *et al.* Titan trace gaseous composition from CIRS at the end of the Cassini-Huygens prime mission. *Icarus* **207,** 461–476 (2010).
17. Teanby, N. A., Irwin, P. G. J., de Kok, R. & Nixon, C. A. Seasonal changes in Titan's polar trace gas abundance observed by Cassini. *Astrophys. J.* **724,** L84–L89 (2010).
18. Vinatier, S. *et al.* Analysis of Cassini/CIRS limb spectra of Titan acquired during the nominal mission I. Hydrocarbons, nitriles and CO$_2$ vertical mixing ratio profiles. *Icarus* **205,** 559–570 (2010).
19. Flasar, F. M. *et al.* Exploring the Saturn system in the thermal infrared: the Composite Infrared Spectrometer. *Space Sci. Rev.* **115,** 169–297 (2004).
20. Irwin, P. *et al.* The NEMESIS planetary atmosphere radiative transfer and retrieval tool. *J. Quant. Spectrosc. Radiat. Transf.* **109,** 1136–1150 (2008).
21. Teanby, N. A. *et al.* Vertical profiles of HCN, HC$_3$N, and C$_2$H$_2$ in Titan's atmosphere derived from Cassini/CIRS data. *Icarus* **186,** 364–384 (2007).
22. Achterberg, R. K., Conrath, B. J., Gierasch, P. J., Flasar, F. M. & Nixon, C. A. Titan's middle-atmospheric temperatures and dynamics observed by the Cassini Composite Infrared Spectrometer. *Icarus* **194,** 263–277 (2008).
23. Achterberg, R. K., Gierasch, P. J., Conrath, B. J., Michael Flasar, F. & Nixon, C. A. Temporal variations of Titan's middle-atmospheric temperatures from 2004 to 2009 observed by Cassini/CIRS. *Icarus* **211,** 686–698 (2011).
24. Lavvas, P. P., Coustenis, A. & Vardavas, I. M. Coupling photochemistry with haze formation in Titan's atmosphere, part II: results and validation with Cassini/Huygens data. *Planet. Space Sci.* **56,** 67–99 (2008).
25. Teanby, N. A., Irwin, P. G. J., de Kok, R. & Nixon, C. A. Mapping Titan's HCN in the far infra-red: implications for photochemistry. *Faraday Discuss.* **147,** 51–64 (2010).
26. Flasar, F. M. & Conrath, B. J. Titan's stratospheric temperatures: a case for dynamical inertia? *Icarus* **85,** 346–354 (1990).
27. West, R. A. *et al.* The evolution of Titan's detached haze layer near equinox in 2009. *Geophys. Res. Lett.* **38,** L06204 (2011).
28. Waite, J. H. *et al.* The process of tholin formation in Titan's upper atmosphere. *Science* **316,** 870–875 (2007).
29. Liang, M.-C., Yung, Y. L. & Shemansky, D. E. Photolytically generated aerosols in the mesosphere and thermosphere of Titan. *Astrophys. J.* **661,** L199–L202 (2007).
30. Krasnopolsky, V. A. A photochemical model of Titan's atmosphere and ionosphere. *Icarus* **201,** 226–256 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by the UK Science and Technology Facilities Council, the Leverhulme Trust and the NASA Cassini mission.

Author Contributions N.A.T. designed the study, performed the radiative transfer analysis and wrote the initial manuscript. P.G.J.I., N.A.T., C.A.N., R.d.K. and S.B.C. developed and maintained the radiative transfer code used for the analysis. S.V. performed independent tests on the results. A.C. performed further checks on the inversion method. All authors contributed to the interpretation of the results, in addition to editing and improving the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.A.T. (n.teanby@bristol.ac.uk).

doi:10.1038/nature11613

Observing the drop of resistance in the flow of a superfluid Fermi gas

David Stadler[1], Sebastian Krinner[1], Jakob Meineke[1], Jean-Philippe Brantut[1] & Tilman Esslinger[1]

The ability of particles to flow with very low resistance is characteristic of superfluid and superconducting states, leading to their discovery in the past century[1,2]. Although measuring the particle flow in liquid helium or superconducting materials is essential to identify superfluidity or superconductivity, no analogous measurement has been performed for superfluids based on ultracold Fermi gases. Here we report direct measurements of the conduction properties of strongly interacting fermions, observing the well-known drop in resistance that is associated with the onset of superfluidity. By varying the depth of the trapping potential in a narrow channel connecting two atomic reservoirs, we observed variations of the atomic current over several orders of magnitude. We related the intrinsic conduction properties to the thermodynamic functions in a model-independent way, by making use of high-resolution *in situ* imaging in combination with current measurements. Our results show that, as in solid-state systems, current and resistance measurements in quantum gases provide a sensitive probe with which to explore many-body physics. Our method is closely analogous to the operation of a solid-state field-effect transistor and could be applied as a probe for optical lattices and disordered systems, paving the way for modelling complex superconducting devices.

Over the past decade, cold atoms have emerged as a many-body system with a uniquely high level of control[3]. Experiments have shown that interacting atomic Fermi gases, analogous to electrons in a solid, can display superfluidity[4]. The equilibrium properties of such gases have been measured with increasing precision[5–8] and the superfluid character of the ground state has been investigated via the response to external perturbations[9] and the direct observation of vortices[10], in the same way as for Bose–Einstein condensates[11–15]. Using new techniques to create and observe directed currents in a closed atomic circuit[16] or between two large reservoirs[17], it is now possible to study the transport properties of mesoscopic systems that are directly analogous to electronic devices[18].

Here we investigated the conduction properties of strongly interacting fermions flowing through a quasi-two-dimensional, multimode channel, which connects two atomic reservoirs. Going beyond our previous work[17], we now obtained full control over the atomic current by tuning a repulsive gate potential in the channel. The gate potential was created by an off-resonant laser beam, as illustrated in Fig. 1. In analogy with an electronic field-effect transistor, this gate potential controls the chemical potential in the channel while keeping the temperature imposed by the reservoirs unchanged. With the gate potential as a control parameter, we measured the current through the channel over a large dynamic range and related it to the observed density in the channel region. This allowed us to observe the onset of superfluid flow of strongly interacting fermions. We compared these measurements to the case of weakly interacting fermions. In our experiment, the current established in the channel was a response to the longitudinal perturbation induced by a difference in chemical potential between the two reservoirs. This is complementary to experiments probing the response of isolated atomic clouds to transverse excitation via rotation[10] or shear[19].

Our experiments were performed with strongly and weakly interacting quantum degenerate gases of fermionic ^{6}Li atoms, equally populating the lowest two hyperfine states. To obtain a strongly interacting gas, the atoms were placed in a homogeneous magnetic field of 834 G where the *s*-wave scattering length diverges and leads to the formation of pairs, while a weakly interacting gas was studied at a field of 475 G (see Methods and Supplementary Information). The atoms were radially confined in the *x*–*z* plane by an optical dipole trap oriented along the *y*-axis with a $1/e^2$ beam radius of 22(1) μm; here and elsewhere, the value in parentheses is the 1-σ error of the last significant digit. Along the *y*-direction, the curvature of the magnetic field yielded a harmonic confinement with a frequency of $\omega_y = 2\pi \times 32(1)$ Hz. To engineer the reservoirs, we split the cloud into two parts using a repulsive laser beam at a wavelength of 532 nm that points along the *x*-direction (beam not shown in Fig. 1). The intensity profile of this beam has a holographically imprinted nodal line along the *y*-axis. As a result, a channel in the *x*–*y* plane was formed, which confined the atoms along the *z*-direction with a centre trap frequency of 2.9 kHz. The gate potential was created by another laser beam at 532 nm that was sent along the *z*-axis onto the channel and had a waist of 18 μm. We refer to the maximum of the repulsive potential created by this beam as the gate potential *U*. Along the *z*-axis, a high-resolution microscope objective was used for *in situ* absorption imaging of the atoms in the channel. The atom number in the reservoirs was measured by absorption imaging along the *x*-direction. By creating an atom number imbalance between the two reservoirs, we created a chemical potential bias that induced a current through the channel[17].

The inset to Fig. 2a presents an example of the time evolution of the relative number imbalance between the two reservoirs, measured for strongly interacting (red) and weakly interacting (blue) fermions, using the same gate potential of $U = 525(50)$ nK. For the strongly interacting gas, an exponential fit yielded a decay time of 0.057(7) s,

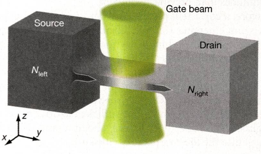


Figure 1 | Principle of the experiment. Two atomic reservoirs (source and drain) are connected by a quasi-two-dimensional conducting channel. An atom number imbalance $N_{\text{left}} > N_{\text{right}}$ between source and drain drives an atom current through the channel, indicated by the arrows. A repulsive laser beam (gate beam) propagating along the *z*-axis is focused on the channel. It creates a repulsive potential with a gaussian envelope and a tunable amplitude. The lighter region in the channel indicates the reduced density due to the repulsive potential.

---

[1] Institute for Quantum Electronics, ETH Zurich, 8093 Zurich, Switzerland.

which is more than one order of magnitude faster than the decay time of 0.70(6) s obtained for the weakly interacting gas.

The reservoirs can be considered to be in quasi-thermal equilibrium during the entire decay, provided this process is sufficiently slow compared to the thermalization dynamics within the reservoirs. Thus we interpret the exponential decay of the imbalance as a resistance measurement through a tunable channel with resistance R. This is analogous to the discharge of a capacitor with a fixed capacity C where the decay time is $\tau = RC$. In our system C is the compressibility of the reservoirs, which remains constant as the gate potential is varied[17]. The natural timescale to which we compare the decay time is provided by ω_y, the frequency of the overall harmonic confinement along the y-axis. Therefore, we defined a dimensionless resistance $r = RC\omega_y$ which is shown in Fig. 2a as a function of the gate potential U. For decreasing gate potential the weakly interacting Fermi gas (blue) shows a decrease of resistance reaching a minimum value of $r \approx 35$ for zero gate potential. For high gate potentials the resistance for both interaction strengths are comparable, yet the strongly interacting gas (red) showed a much faster drop of resistance below 0.7 μK. At a gate potential of 0.23(2) μK the resistance differed by a factor of about 25 from the weakly interacting gas. As r approaches unity (below 0.23 μK) the decay time τ became equal to the timescale of the internal dynamics of the reservoirs, set by the trap frequency along the y-direction. In this regime, we cannot interpret our strongly interacting data sets in terms of a resistance measurement because the reservoirs do not remain in thermal equilibrium at each point in time, that is, the resistance drops below our measurement capabilities. This gives rise to deviations from the exponential decay.

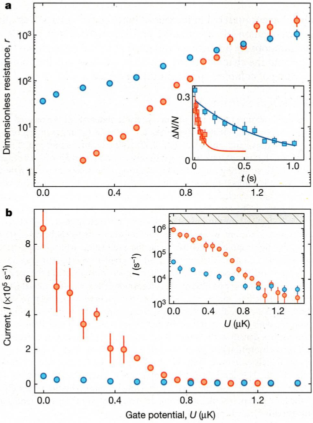


Figure 2 | Conduction properties through the channel. Red and blue data points correspond to the strongly and weakly interacting gas, respectively. **a**, Dimensionless resistance r as a function of gate potential U. The data points shown are those for which the decay is exponential. The inset to **a** shows a decay of the relative atom number imbalance between source and drain as a function of time with a gate potential $U = 525(50)$ nK, where N is the total number of atoms and ΔN is the difference in atom number. The solid lines are exponential fits with fixed offset of 0.04 for the red curve to account for a small remaining imbalance in the reservoirs. **b**, Atom current as a function of the gate potential U. A large increase of the current appears for the strongly interacting gas below $U \approx 0.7$ μK. The inset to **b** shows the atom current in logarithmic scale. The shaded region indicates the maximum current allowed by the internal dynamics of the reservoirs (see main text). The error bars show the statistical errors.

In addition to the resistance, we also estimated the current through the channel using a linear fit to the initial part of the decay (see Methods). This measurement does not rely on the thermalization of the reservoirs and thus can also be applied to cases where the reservoirs are not fully in quasi-thermal equilibrium. Figure 2b shows the current I as a function of the gate potential for the strongly interacting gas (red) and the weakly interacting gas (blue). Unlike the weakly interacting gas, the strongly interacting gas showed a fast increase of the current below 0.7 μK. For the lowest gate potentials the current was limited by the conservation of energy. The limit was reached when the potential energy introduced by the initial imbalance was fully converted into kinetic energy, as for example in undamped dipole oscillations. It is represented by the shaded region in the inset to Fig. 2b, where we show the current in logarithmic scale. Remarkably, the observed current was very close to that limit, meaning that the strongly interacting Fermi gas flowed as if there were no constriction or gate potential at all. This is the expected behaviour of a superfluid.

Although the current depends on the atomic density in the channel, the transport properties are characterized in a density-independent way by the drift velocity. To extract this quantity, we first used high-resolution *in situ* imaging to measure the atomic line density n_{line} in the channel. The measured line density as a function of the gate potential is shown in the inset to Fig. 3. As expected from its higher compressibility[8,20], the strongly interacting gas reached larger line densities. For each value of the gate potential, we then divided the measured current by the corresponding line density, yielding the drift velocity.

The drift velocities as a function of gate potential are presented in Fig. 3. The drift velocity for the weakly interacting gas showed almost no variations. In contrast, the drift velocity for the strongly interacting gas increased significantly below $U = 0.7$ μK. This demonstrates that the large increase of the current, seen in Fig. 2, was not simply caused by the higher density of the strongly interacting gas in the channel and reveals a change in the nature of the transport process. It cannot be explained by a transition of the gas from ballistic to classical hydrodynamic behaviour because even for large gate potentials the mean free

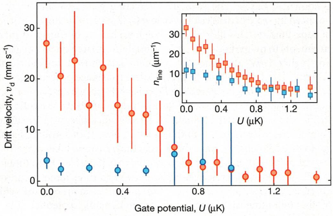


Figure 3 | Density-independent conduction properties through the channel. Red and blue data points correspond to strongly interacting and weakly interacting atoms, respectively. Drift velocity is plotted as a function of gate potential. The points corresponding to the three highest values of the gate potential are omitted in the weakly interacting case because the density is almost zero. The inset shows line density n_{line} measured *in situ* in the channel as a function of gate potential (see Methods). The error bars represent statistical errors (one standard deviation).

path remains well below the channel size. On the other hand, at low gate potentials, that is, at low T/T_F, where T_F is the Fermi temperature, Pauli blocking of interparticle collisions in a normal gas should restore the ballistic behaviour[21–23] (see Methods). This is in contrast to our observations, supporting our superfluidity interpretation. Whether another mechanism can lead to better conduction properties than that of a perfect ballistic conductor is unclear.

It was instructive to compare the drift velocity to characteristic velocities of the superfluid flow. A Landau-type critical velocity provides a value which is expected to be of the order of the Fermi velocity for strongly interacting fermions[4]. We estimated the Fermi velocity in the channel from the column density at zero gate potential, which gives about 50 mm s^{-1}, twice as large as the measured drift velocity. Additionally, the creation of vortices in the fluid provided a lower critical velocity, giving rise to a finite resistance. This velocity can be roughly estimated from the channel geometry and the healing length using energetic arguments[16], and yielded about 5 mm s^{-1}. The observed drift velocity was significantly larger than this critical value. This would explain the low but finite resistance observed even in the superfluid state, where the decay of the number imbalance is fast but remains exponential.

We next related the conduction properties to a thermodynamic parameter by replacing the gate potential scale, which is specific to our system, by the thermodynamic potential. To this end, we used the high-resolution images of the gas in the channel, which gave us access to the equation of state[6–8]. The gas in the channel is in the crossover regime between two and three dimensions, where the equation of state naturally relates the column density n_{col} to the chemical potential[24,25] (see Methods). From the *in situ* absorption images of the channel for different gate potentials we obtained $n_{\mathrm{col}}(U)$ at fixed temperature, which is imposed by the reservoirs. Integrating this relation over the known variations of the gate potential yielded the thermodynamic potential $\Omega(U)=\int_U^\infty n_{\mathrm{col}}(V)\mathrm{d}V$ which would be equal to the pressure in a purely two-dimensional gas. We normalized Ω by the pressure of a two-dimensional ideal Fermi gas at zero temperature $\Omega_0=\pi\hbar^2 n_{\mathrm{col}}^2/m$ and obtained a model-independent thermodynamic scale, analogous to the three-dimensional situation discussed in ref. 8. This allowed us to convert the gate potential into a thermodynamic quantity, even though the gas in the channel was not expected to be in the universal regime[4] owing to the strong confinement[26], where most of the thermometry techniques cannot be applied directly[5,8].

The drift velocity as a function of reduced thermodynamic potential is shown in Fig. 4. The strongly interacting gas (red) show a pronounced increase of drift velocity below $\Omega/\Omega_0\approx 1$, indicating the onset of superfluidity. This illustrates the high sensitivity of transport measurements to many-body effects in strongly correlated quantum gases. For higher Ω/Ω_0 the blue and red data sets show a constant drift velocity. The inset to Fig. 4b presents the resistance as a function of Ω/Ω_0 for the strongly interacting Fermi gas. Here, we observed a very rapid decrease of the resistance for low values of Ω/Ω_0. We interpret this as the counterpart of the drop of resistance observed in superconductors. Measurements of the equation of state of a unitary Fermi gas in three dimensions have shown that the transition takes place for a critical reduced thermodynamic potential of 0.55 (ref. 8). Even though our channel is in the crossover between two and three dimensions, we observed the change in the conduction properties at around the value of the two-dimensional reduced thermodynamic potential (black dashed lines in Fig. 4).

Our experimental geometry is reminiscent of weak links in superconductors[1] and the experiment probes transport in a channel that is long compared to the coherence length. The coherence length of a strongly interacting superfluid is of the order of the interparticle spacing[4], which is below a few micrometres in the channel and smaller elsewhere. The length and energy scales of our experiment mean that we operate in a dissipative regime complementary to the coherent tunnelling encountered in Josephson junctions[27,28]. Our set-up allows the investigation of superfluidity and supercurrents in a variety of configurations by projecting a designed potential through the microscope onto the channel[29]. This opens the way towards the cold-atom modelling of complex, superconducting devices.

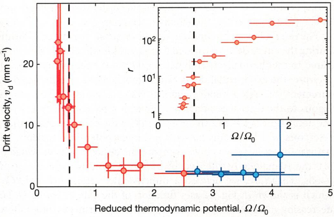


Figure 4 | Conduction properties as a function of thermodynamic potential. Drift velocity is plotted as a function of the reduced thermodynamic potential Ω/Ω_0 for the strongly interacting (red) and weakly interacting (blue) Fermi gas. The inset shows dimensionless resistance as a function of Ω/Ω_0 in logarithmic scale for the strongly interacting gas, showing the drop of resistance. The dashed black lines at $\Omega/\Omega_0=0.55$ indicate the position where the superfluid transition in three dimensions occurs. Error bars represent statistical errors (one standard deviation).

METHODS SUMMARY

A balanced mixture of the two lowest hyperfine states of ^{6}Li is prepared by all-optical evaporation. Final temperatures are $\lesssim 0.1T_F$ (strongly interacting gas, 6.7×10^4 atoms) and approximately $0.3T_F$ (weakly interacting gas, 4.5×10^4 atoms). For the strongly interacting gas the evaporation is performed at a magnetic field of 795 G (scattering length $3{,}500a_0$, where a_0 is the Bohr radius), then the field is adiabatically ramped to 834 G, at the *s*-wave Feshbach resonance. The weakly interacting gas is cooled at 300 G, then the field is ramped to 475 G (scattering length $-100a_0$). The trap frequency along the *y*-axis is $\omega_y=2\pi\times32(1)\ \mathrm{s}^{-1}$ and $\omega_y=2\pi\times25(1)\ \mathrm{s}^{-1}$ for the strongly and weakly interacting gases, respectively. To induce an atom current, we create a number imbalance between the two reservoirs by shifting the trapping potential along the *y*-direction with a magnetic field gradient of 0.25 G cm^{-1}. After switching off the gradient within 10 ms, we monitor the decay of the number imbalance. The number imbalance and the total atom number are obtained from absorption images along the *x*-axis. For all data, we fit a line to the first four points of a measured decay curve of the relative number imbalance. We define the current as the fitted slope times half the total number of atoms in both reservoirs at equilibrium. To measure the column density n_{col}, as well as the line density at the centre of the channel for different gate potentials, we take *in situ* absorption images of the channel through the high-resolution microscope in the absence of current. We apply light pulses of 5 μs and a saturation of about 0.1. The local density approximation gives the equation of state[7] $n_{\mathrm{col}}(\mu_0-V)$, where μ_0 is the chemical potential imposed by the reservoirs and V is the local gate potential. Integrating this equation over the gate potential leads to the thermodynamic potential.

Full Methods and any associated references are available in the online version of the paper.

Received 6 June; accepted 20 September 2012.

1. Leggett, A. J. *Quantum Liquids: Bose Einstein Condensation and Cooper Pairing in Condensed-Matter Systems* (Oxford University Press, 2006).
2. van Delft, D. & Kes, P. The discovery of superconductivity. *Phys. Today* **63,** 38–43 (2010).
3. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80,** 885–964 (2008).
4. Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of ultracold atomic Fermi gases. *Rev. Mod. Phys.* **80,** 1215–1274 (2008).
5. Luo, L., Clancy, B., Joseph, J., Kinast, J. & Thomas, J. E. Measurement of the entropy and critical temperature of a strongly interacting Fermi gas. *Phys. Rev. Lett.* **98,** 080402 (2007).

6. Horikoshi, M., Nakajima, S., Ueda, M. & Mukaiyama, T. Measurement of universal thermodynamic functions for a unitary Fermi gas. *Science* **327,** 442–445 (2010).
7. Nascimbène, S., Navon, N., Jiang, K. J., Chevy, F. & Salomon, C. Exploring the thermodynamics of a universal Fermi gas. *Nature* **463,** 1057–1060 (2010).
8. Ku, M. J. H., Sommer, A. T., Cheuk, L. W. & Zwierlein, M. W. Revealing the superfluid lambda transition in the universal thermodynamics of a unitary Fermi gas. *Science* **335,** 563–567 (2012).
9. Miller, D. E. *et al.* Critical velocity for superfluid flow across the BEC-BCS crossover. *Phys. Rev. Lett.* **99,** 070402 (2007).
10. Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunck, C. H. & Ketterle, W. Vortices and superfluidity in a strongly interacting Fermi gas. *Nature* **435,** 1047–1051 (2005).
11. Madison, K. W., Chevy, F., Wohlleben, W. & Dalibard, J. Vortex formation in a stirred Bose-Einstein condensate. *Phys. Rev. Lett.* **84,** 806–809 (2000).
12. Matthews, M. R. *et al.* Vortices in a Bose-Einstein condensate. *Phys. Rev. Lett.* **83,** 2498–2501 (1999).
13. Raman, C. *et al.* Evidence for a critical velocity in a Bose-Einstein condensed gas. *Phys. Rev. Lett.* **83,** 2502–2505 (1999).
14. Burger, S. *et al.* Superfluid and dissipative dynamics of a Bose-Einstein condensate in a periodic optical potential. *Phys. Rev. Lett.* **86,** 4447–4450 (2001).
15. Amo, A. *et al.* Superfluidity of polaritons in semiconductor microcavities. *Nature Phys.* **5,** 805–810 (2009).
16. Ramanathan, A. *et al.* Superflow in a toroidal Bose-Einstein condensate: an atom circuit with a tunable weak link. *Phys. Rev. Lett.* **106,** 130401 (2011).
17. Brantut, J.-P., Meineke, J., Stadler, D., Krinner, S. & Esslinger, T. Conduction of ultracold Fermions through a mesoscopic channel. *Science* **337,** 1069–1071 (2012).
18. Seaman, B. T., Krämer, M., Anderson, D. Z. & Holland, M. J. Atomtronics: Ultracoldatom analogs of electronic devices. *Phys. Rev. A* **75,** 023615 (2007).
19. Cao, C. *et al.* Universal quantum viscosity in a unitary Fermi gas. *Science* **331,** 58–61 (2011).
20. Bartenstein, M. *et al.* Crossover from a molecular Bose-Einstein condensate to a degenerate Fermi gas. *Phys. Rev. Lett.* **92,** 120401 (2004).
21. Sommer, A., Ku, M., Roati, G. & Zwierlein, M. W. Universal spin transport in a strongly interacting fermi gas. *Nature* **472,** 201–204 (2011).
22. Enss, T., Haussmann, R. & Zwerger, W. Viscosity and scale invariance in the unitary Fermi gas. *Ann. Phys.* **326,** 770–796 (2011).
23. Bruun, G. M. Shear viscosity and spin-diffusion coefficient of a twodimensional Fermi gas. *Phys. Rev. A* **85,** 013636 (2012).
24. Orel, A. A., Dyke, P., Delehaye, M., Vale, C. J. & Hu, H. Density distribution of a trapped two-dimensional strongly interacting Fermi gas. *N. J. Phys.* **13,** 113032 (2011).
25. Dyke, P. *et al.* Crossover from 2D to 3D in a weakly interacting Fermi gas. *Phys. Rev. Lett.* **106,** 105304 (2011).
26. Petrov, D. S. & Shlyapnikov, G. V. Interatomic collisions in a tightly confined Bose gas. *Phys. Rev. A* **64,** 012706 (2001).
27. Albiez, M. *et al.* Direct observation of tunneling and nonlinear self-trapping in a single bosonic Josephson junction. *Phys. Rev. Lett.* **95,** 010402 (2005).
28. LeBlanc, L. J. *et al.* Dynamics of a tunable superfluid junction. *Phys. Rev. Lett.* **106,** 025302 (2011).
29. Zimmermann, B., Müller, T., Meineke, J., Esslinger, T. & Moritz, H. High-resolution imaging of ultracold fermions in microscopically tailored optical potentials. *N. J. Phys.* **13,** 043007 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge discussions with W. Zwerger, A. Georges, C. Kollath, C. Grenier, M. Sigrist, G. Blatter and G. Bruun. We thank L. Tarruell, T. Donner and H. Moritz for their careful reading of the manuscript. We acknowledge financing from NCCR MaNEP and QSIT, ERC project SQMS, FP7 project NAME-QUAM and ETHZ. J.-P.B. acknowledges support from EU through a Marie Curie Fellowship.

Author Contributions All authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-P.B. (brantutj@phys.ethz.ch) or T.E. (esslinger@phys.ethz.ch).

LETTER

doi:10.1038/nature11609

A canonical stability–elasticity relationship verified for one million face-centred-cubic structures

Sascha B. Maisel[1], Michaela Höfler[1] & Stefan Müller[1]

Any thermodynamically stable or metastable phase corresponds to a local minimum of a potentially very complicated energy landscape. But however complex the crystal might be, this energy landscape is of parabolic shape near its minima. Roughly speaking, the depth of this energy well with respect to some reference level determines the thermodynamic stability of the system, and the steepness of the parabola near its minimum determines the system's elastic properties. Although changing alloying elements and their concentrations in a given material to enhance certain properties dates back to the Bronze Age[1,2], the systematic search for desirable properties in metastable atomic configurations at a fixed stoichiometry is a very recent tool in materials design[3]. Here we demonstrate, using first-principles studies of four binary alloy systems, that the elastic properties of face-centred-cubic intermetallic compounds obey certain rules. We reach two conclusions based on calculations on a huge subset of the face-centred-cubic configuration space. First, the stiffness and the heat of formation are negatively correlated with a nearly constant Spearman correlation[4] for all concentrations. Second, the averaged stiffness of metastable configurations at a fixed concentration decays linearly with their distance to the ground-state line (the phase diagram of an alloy at zero Kelvin). We hope that our methods will help to simplify the quest for new materials with optimal properties from the vast configuration space available.

To the best of our knowledge, the only publication that successfully presented a full configurational optimization as a tool to systematically search for metastable phases harder than their respective ground states was by Yuge[3]. The author reported metastable boron-carbon nitride configurations with bulk moduli almost as high as that of pure diamond. Because both heat of formation and bulk modulus are functions of the carbon concentration, the author concluded that a very natural correlation existed between heat of formation and elastic properties—ultimately, this was a result of the dependence of both quantities on the concentration. Does any correlation remain if we restrain ourselves to one specific concentration? We shall make two claims and verify them on a subset of more than one million face-centred-cubic (f.c.c.) structures. These structures have been sampled from the configuration space of four characteristic binary f.c.c. intermetallics, all of them technologically relevant materials for a multitude of applications: nickel-aluminium near the Ni_3Al stoichiometry (the primary precipitating phase constituting the γ'-phase in nickel-based superalloys), Ni-Ta (another vital precipitor in Ni-rich alloys), Cu-Al (a very common light-weight stainless alloy) and nickel-rich Ni-W (a high-temperature alloy for extreme conditions). These four are characteristic of the family of the f.c.c. binary alloys. However, our approach could readily be used to test both claims in systems with covalent[3] or ionic bonds, or in systems that do not crystallize on the f.c.c. lattice.

Claim I is as follows: for any subset X_x of the f.c.c. configuration space with constant concentration $x = \text{const}$, the heat of formation $\Delta H(\sigma)$ of structure σ and its elastic stiffness $\overline{c_{ii}}$ are negatively correlated. Specifically, for all structures $\sigma \in X_x$, both averaged tetragonal stiffness $\overline{c_{11}}$ and shear stiffness $\overline{c_{44}}$ are almost monotonic functions of the enthalpy excess $\beta = (\Delta H(\gamma) - \Delta H(\sigma))/(\Delta H(\gamma))$ of structure σ with respect to the structure γ with lowest ΔH at the respective concentration. The dimensionless enthalpy excess β is a very useful quantity in our case, where several alloys have been studied and the actual values for ΔH vary (see Supplementary Information) between 0.43 meV (γ'-Ni_3Al) and 10 meV (for CuAl) per atom. Hence, expressing the elastic properties as functions of β instead of ΔH allows for a direct comparison of the correlation in the four investigated intermetallics. The claimed correlation can be quantified via the Spearman correlation[4] coefficient, ρ. This quantity takes on values between $-1 \leq \rho \leq 1$, where a Spearman correlation of $\rho(a,\, b) = 1$ implies that $a(b)$ is a strictly monotonic increasing function of b with $\partial a(b)/\partial b > 0$ and vice versa. On the other hand, a Spearman correlation of $\rho(a,\, b) = -1$ implies a strictly monotonic decrease (anti-correlation) with $\partial a(b)/\partial b < 0$. Using Spearman's ρ, claim I can be condensed into a single equation: $\rho(\beta, \overline{c_{ii}}) < 0$.

From a mathematician's point of view, claim I states that the depth of the potential well is not independent of the derivatives of the energy parabolas near its local minima. The actual magnitude of $\rho(\beta, \overline{c_{11}})$ and $\rho(\beta, \overline{c_{44}})$ will quantify the degree of correlation, the quality of monotony and (together with the number of structures in the subset X_x) the likeliness of outliers. From a materials scientist's point of view, a successful proof of claim I will immediately raise a plethora of questions about the exact form of the correlation. Consider the problem of manufacturing a material with certain mechanical properties—a task required since the dawn of metal casting[5]. One would be bound to ask if the correlation severely restricts the usefulness of dipping into non-equilibrium states when trying to grow a very hard crystal. Also, the nature of the elasticity–stability relationship could be susceptible to exploitation if the task at hand is to soften a material. This has been successfully performed in several commercial titanium alloys[6–8], which have been softened by inducing a metastable body-centred-cubic (b.c.c.) phase which is softer than the ground state. This effectively reduces the stiffness to the point where the alloy is soft enough to serve as a material for hip implants without damaging the adjacent bones[6]. This typically involves a lattice change from structures based on hexagonal close packing to b.c.c.-based B2-type structures, and thus a change of the underlying lattice[8]. Our findings here support the suggestion that a similar softening could be achieved without causing lattice changes.

To verify claim I, we have exhaustively enumerated at least 1.7×10^5 f.c.c.-based configurations for each of the four alloys up to a unit cell size of 21 atoms. The enumeration has been performed using an algorithm

Table 1 | Cross-validation scores S_{cv} for all 12 cluster expansions

Score	Alloy			
	γ'-NiAl	Cu-rich CuAl	NiTa	NiW
$S_{cv}(\Delta H)$ (eV)	0.94×10^{-3}	0.26×10^{-3}	2.52×10^{-3}	5.92×10^{-3}
$S_{cv}(\overline{c_{11}})$ (GPa)	1.71	5.71	2.34	3.21
$S_{cv}(\overline{c_{44}})$ (GPa)	1.51	0.62	1.09	2.25

Scores are quantitative measures of the predictive power of these expansions.

---

[1] Hamburg University of Technology, Institute of Advanced Ceramics, Denickestraße 15, 21073 Hamburg, Germany.

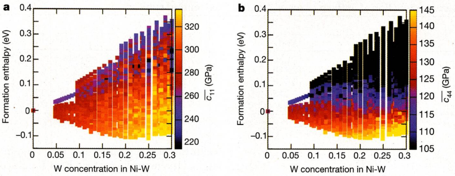


Figure 1 | Ground-state diagrams of binary Ni-W. a, b, Colour coding shows $\overline{c_{11}}$(**a**) and $\overline{c_{44}}$(**b**). Each point in the diagrams corresponds to one specific geometric arrangement σ of atoms with given tungsten concentration on the lattice sites of a face-centred cubic-lattice. Their respective formation enthalpies and elastic properties are deduced from the cluster-expansions (equations (1)-(3)) and indicated on the y axis and the colour-coding, respectively. Similar diagrams for binary Ni-Ta, Cu-rich Cu-Al and γ'-Ni-Al and can be found in Supplementary Information.

presented in refs 9 and 10, and yields a grand total of 1,405,970 structures. Then, density functional theory has been applied to a representative subset of configurations for each of the four systems. These calculations have been performed using the Vienna Ab-Initio Simulation package (VASP)[11,12]; computational details and input parameters can be found in Methods. Using the data sets of density functional energies, an extrapolation of the energies of the remaining enumerated structures has been performed using the cluster expansion technique, as realized in our UNCLE package[13]. The cluster expansion approach uses an analytical decomposition of the many-body interaction inherent in any crystal. Any observable that is strictly a function of the atomic configuration σ can be expanded into a sum over many-body correlation functions Π_F times some expansion coefficients J_F, which has been proven[14,15]. Because both the symmetrically averaged elastic moduli $\overline{c_{ii}}$ and the heat of formation ΔH per atom are functions of the atomic structure σ, the cluster expansion method is ideally suited to expand both the mechanical properties and the energetics:

$$\overline{c_{11}}(\sigma)=\sum_F J_F^{(11)}\Pi_F(\sigma) \qquad (1)$$

$$\overline{c_{44}}(\sigma)=\sum_F J_F^{(44)}\Pi_F(\sigma) \qquad (2)$$

$$\Delta H(\sigma)=-x_{\mathrm{A}}(\sigma)E_{\mathrm{A}}-x_{\mathrm{B}}(\sigma)E_{\mathrm{B}}+\sum_F J_F^{(E)}\Pi_F(\sigma) \qquad (3)$$

where E_{A} and E_{B} are the free energies per atom of the pure elements, of which structure σ contains a fraction x_{A} of component A and x_{B} of component B. Using equations (1)–(3), a grand total of 12 expansions have been carried out. The average deviation calculated using a cross-validation score S_{cv} for each expansion is given in Table 1. This score S_{cv} is a measure of the predictive power of a cluster expansion[13]. Figure 1b shows a ground-state diagram of binary Ni-W, where the averaged shear stiffness $\overline{c_{44}}$ has been colour-coded into the diagram. The very stiff structures depicted in yellowish colours are predominantly found near the ground-state line, while the very soft structures are realized at high formation enthalpies and therefore at high enthalpy excess β. Similar diagrams can be found for both the tetragonal modulus and the shear stiffness for all four intermetallic systems; see Supplementary Information.

The data set used to generate Fig. 1 can be used to verify claim I by directly calculating Spearman's coefficient $\rho(\beta,\overline{c_{ii}})$ for a large sample size (Fig. 2). As claimed, both $\rho(\beta,\overline{c_{11}})$ and $\rho(\beta,\overline{c_{44}})$ are strictly negative for all systems and all compositions. Hence, the stiffness is a (quasi) strictly decreasing function of the enthalpy excess β of the structures σ in X_x over which $\overline{c_{ii}}(\sigma)$ is sampled. We note that the correlation coefficient is specific for the material and the respective elastic property, and does not vary much with concentration. This behaviour appears to be generic, which suggests that Spearman's ρ is a quantity characteristic for each ordered phase in intermetallic alloys. As Spearman's ρ only quantifies the degree of correlation without actually fixing the functional relation, this immediately prompts the formulation of another claim.

Claim II is as follows: let $X_{\beta,x}$ be the set of all f.c.c. structures with enthalpy excess β and $x(\sigma)=x$. Let $\langle\overline{c_{44}}(\sigma)\rangle$ be the averaged shear modulus averaged over the set $X_{\beta,x}$. Then, the harmonic (Voigt-type) average

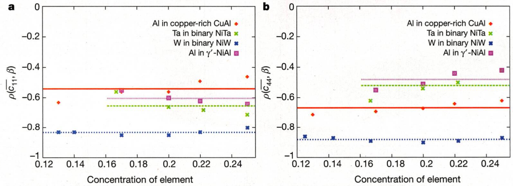


Figure 2 | Spearman's coefficients $\rho(\beta,\overline{c_{11}})$ and $\rho(\beta,\overline{c_{44}})$ for the correlation between stiffness $\overline{c_{ii}}$ and enthalpy excess β. a, $\rho(\beta,\overline{c_{11}})$; **b,** $\rho(\beta,\overline{c_{44}})$. The x axis shows concentrations of Al in copper-rich Cu-Al (red crosses), Ta in binary Ni-Ta (green crosses), W in binary Ni-W, and Al in γ'-Ni-Al. The correlation is strictly negative (anticorrelation) for all systems and compositions. Note that Spearman's ρ is nearly constant for a given material and choice of i. The constant fits have been determined by generating ten sets of interaction coefficients in equations (1)–(3) for each system with subsequent calculation of the arithmetic mean.

Table 2 | Decay of stiffness as a function of enthalpy excess β for five face-centred cubic alloys

Structures and fits	Alloy				
	γ'-$Ni_{75}Al_{25}$	$Ni_{80}Ta_{20}$	$Cu_{83.4}Al_{16.6}$	$Cu_{80}Al_{20}$	$Ni_{80}W_{20}$
Structures	29,649	11,500	1,474	8,763	9,753
Fit residual $\langle\overline{c_{11}}\rangle$ (GPa)	0.841	2.478	4.352	2.661	2.79
Fit residual $\langle\overline{c_{44}}\rangle$ (GPa)	0.433	0.958	0.568	0.470	0.485
$\partial\langle\overline{c_{11}}\rangle/\partial\beta$ (GPa)	-76.76 ± 0.07	-34.56 ± 0.08	-50.04 ± 0.69	-38.92 ± 0.17	-24.72 ± 0.06
$\partial\langle\overline{c_{44}}\rangle/\partial\beta$ (GPa)	-16.09 ± 0.04	-61.00 ± 0.05	-19.56 ± 0.17	-25.15 ± 0.05	-19.05 ± 0.03

Stiffness is given by $\partial\langle\overline{c_{ii}}\rangle/\partial\beta$, and corresponds to the slopes of the linear fits in Fig. 4. Harmonic averaging has been used for the thermal averaging denoted by $\langle\rangle$.

$$\langle\overline{c_{44}}\rangle(\beta,x) = \frac{N(\beta,x)}{\sum \frac{1}{\overline{c_{44}}(\sigma)}} \tag{4}$$

over N different structures will decrease approximately linearly as a function of the enthalpy excess β in the vicinity of the ground-state line $0 \leq \beta \ll 1$. A similar functional relation will hold for the average of $\overline{c_{11}}(\beta,x)$. The thermal averaging denoted by $\langle\rangle$ is crucial, because at finite temperatures, all structures with equal enthalpy excess β $(\rho) =$ const are in coequal competition.

To validate claim II, we have analysed $\langle\overline{c_{11}}\rangle$ (β, x) and $\langle\overline{c_{44}}\rangle(\beta, x)$ for five different f.c.c. systems (Table 2). The five data sets used for the averaging process have been generated from the previous cluster expansions. Thus, the five systems correspond to five columns at specific concentrations in the diagrams of Fig. 1 and Supplementary Fig. 1. To illustrate this, Fig. 3 shows $\overline{c_{44}}$ (β, x) for $Ni_{80}W_{20}$ and Ni_3Al before the thermal averaging.

Figure 4 shows the results of averaging over all structures at constant β according to equation (4) for the five compositions in Table 2. According to claim II, the decay of $\overline{c_{11}}(\beta)$ and $\overline{c_{44}}(\beta)$ as a function of the enthalpy excess β is roughly linear for each of the five systems and Fig. 4 clearly confirms that. As a quantitative confirmation, fit residuals are supplied in Table 2, and are typically only a few GPa.

From an application-oriented perspective, the successful verification of claim II allows us to predict the odds of finding the desired properties in structures with given enthalpy excess β. Touching again on the example of the Ti-based hip joints mentioned above—the shape of Fig. 4 suggests that more in-depth research on the relationships presented here will allow materials to be tuned to the desired stiffness in similar circumstances, where an application requires a material with exactly the right elastic properties. The verification of claim II also establishes that looking for very stiff metastable phases should always commence at the ground-state line. With increasing $\beta > 0$, it will be less and less rewarding to search for atomic configurations with a high stiffness—the steepest parabolas and hence the stiffest specimens are encountered in the stable and nearly stable phases.

Although all alloy compositions show a more or less linear decay of $\langle\overline{c_{11}}\rangle(\beta,x)$ and $\langle\overline{c_{44}}\rangle(\beta,x)$, the slopes $\partial\overline{c_{ii}}/\partial\beta$ vary significantly from material to material. Most notably, the shear stiffness $\overline{c_{44}}$ of the binary Ni-Ta alloy decreases rapidly away from the ground-state line: the $\partial\overline{c_{44}}/\partial\beta$ values for Ni-Ta are the highest of all the materials we investigated. From a materials scientist's perspective, this corresponds to a system which loses its mechanical stability quickly, if the crystal structure is not near its thermodynamic equilibrium. Assuming that the proposed linear behaviour is generic, systematic tabulation of $\partial\overline{c_{ii}}/\partial\beta$ for different systems appears to be a worthwhile endeavour for both crystallographic and application-driven purposes.

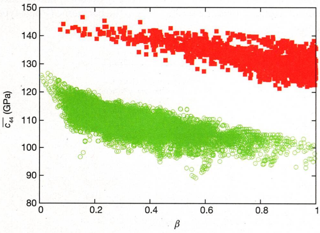


Figure 3 | $\overline{c_{44}}$ as a function of the enthalpy excess β before averaging over all structures at constant β. Each coloured dot corresponds to one real-space structure. All structures with fixed enthalpy excess are in coequal competition. Roughly 1,500 structures out of a $Ni_{80}W_{20}$ subset (red squares) of the f.c.c. configuration space are shown, together with 35,000 structures out of a $Ni_{75}Al_{25}$ subset (green circles) of that configuration space.

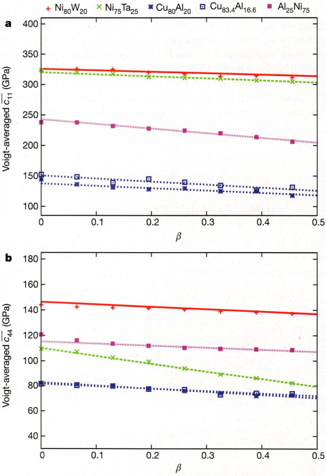


Figure 4 | Averaged stiffnesses using harmonic (Voigt-type) averaging for four different systems at five different compositions. a, $\langle\overline{c_{11}}\rangle(\beta,x)$; **b,** $\langle\overline{c_{44}}\rangle(\beta,x)$. Key shows systems and compositions examined. Note that a change of composition from the $Cu_{80}Al_{20}$ alloy to the stiffer $Cu_{83.4}Al_{16.6}$ merely leads to a nearly parallel shift of the stiffness curves, without drastically changing the functional relation or the slopes.

In this work, we have offered a systematic large-scale analysis of the elasticity–enthalpy dependence for f.c.c.-based metal–metal alloys only. However, the methods that we have used to demonstrate these correlations and relationships could be used to investigate whether these claims also hold for other systems. Several indications of a configuration dependence of, for example, bulk moduli have been given in the literature for practically all classes of materials—not only intermetallic alloys[16,17], but also glasses[18] and even carbides[3]. Further work is required to verify or falsify our findings for other base lattices and different classes of materials.

METHODS SUMMARY

The four Hamiltonians have been generated using the cluster-expansion approach. The method is introduced elsewhere[14,15] and the actual implementation used is the UNCLE package[13]. The basis set of the expansion consists of real-space correlation functions. The expansion coefficients of the four different cluster expansions have been obtained by fitting to density functional theory input calculations, employing an evolutionary approach using genetic algorithms[19,20]. The first-principles input required by the genetic algorithms has been provided using the VASP package[11,12]. The exchange-correlation functional of the DFT calculations was described within the general gradient approximation as parameterized in ref. 21. Specifically, the PBE-parameterized PAW-GGA-potentials[21,22] as supplied with VASP have been employed. All structures have been geometrically fully relaxed. Integration within the Brillouin zone has been performed using Γ-centred k-meshes with up to $23 \times 23 \times 23$ grid points depending on supercell size, employing the usual Methfessel-Paxton smearing[23] for the relaxations and the Blöchl-corrected tetrahedron method[24] to determine the energies. After fully relaxed geometries have been obtained, the elastic properties were determined by applying the finite distortions formalism of VASP using the four-displacement scheme ($\text{NFREE} = 4$) at a spacing of 0.018 Å. As this procedure requires absolute convergence of the stress tensor, convergence tests of the plane-wave energy cut-offs had to be performed for all relaxations and distortions. At their conclusion, increased cut-offs of 420 eV were chosen for all systems. The concentration ranges of enumerated structures using the algorithm introduced in refs 9 and 10 vary from system to system (Supplementary Fig. 5). This is because we took particular care to have all structures in all systems and at all compositions reside on an f.c.c.-based lattice in one single phase, since for some systems, especially NiTa (ref. 25), the elastic properties vary significantly between neighbouring phases—an effect that is potentially stronger than the decay of stiffness away from equilibrium.

Full Methods and any associated references are available in the online version of the paper.

Received 28 June; accepted 11 September 2012.
Published online 21 November 2012.

1. Craddock, P. T. The composition of the copper alloys used by the Greek, Etruscan and Roman Civilisations. 1. The Greeks before the Archaic period. *J. Archaeol. Sci.* **3,** 93–113 (1976).
2. Craddock, P. T. The composition of the copper alloys used by the Greek, Etruscan and Roman civilisations: 2. The Archaic, Classical and Hellenistic Greeks. *J. Archaeol. Sci.* **4,** 103–123 (1977).
3. Yuge, K. Prediction of superhard cubic boron-carbon nitride through first principles. *J. Phys. Condens. Matter* **21,** 415403 (2009).
4. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **15,** 72–101 (1904).
5. Wertime, T. A. The beginnings of metallurgy: a new look. *Science* **182,** 875–887 (1973).
6. Raabe, D. *et al.* Theory-guided bottom-up design of β-titanium alloys as biomaterials based on first principles calculations: theory and experiments. *Acta Mater.* **55,** 4475–4487 (2007).
7. Saito, T. *et al.* Multifunctional alloys obtained via a dislocation-free plastic deformation mechanism. *Science* **300,** 464–467 (2003).
8. Holec, D., Friák, M., Dlouhy, A. I. & Neugebauer, J. Ab initio study of pressure stabilized NiTi allotropes: pressure-induced transformations and hysteresis loops. *Phys. Rev. B* **84,** 224119 (2012).
9. Hart, G. L. W. & Forcade, R. W. Algorithm for generating derivative structures. *Phys. Rev. B* **77,** 224115 (2008).
10. Hart, G. L. W. Where are nature's missing structures? *Nature Mater.* **6,** 941–945 (2007).
11. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54,** 11169–11186 (1996).
12. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6,** 15–50 (1996).
13. Lerch, D. *et al.* UNCLE: a code for constructing cluster expansions for arbitrary lattices with minimal user-input. *Model. Simul. Mater. Sci. Eng.* **17,** 055003 (2009).
14. Sanchez, J. M. & de Fontaine, D. in *Structure and Bonding in Crystals* Vol. 2, 117 (Academic, 1981).
15. Sanchez, J. M., Ducastelle, F. & Gratias, D. Generalized cluster expansion of multicomponent systems. *Physica A* **128,** 334–350 (1984).
16. Liu, J. Z., van de Walle, A., Ghosh, G. & Asta, M. Structure, energetics, and mechanical stability of Fe-Cu bcc alloys from first-principles calculations. *Phys. Rev. B* **72,** 144109 (2005).
17. Zhu, L.-F. *et al.* First-principles study of the thermodynamic and elastic properties of eutectic FeTi alloys. *Acta Mater.* **60,** 1594–1602 (2012).
18. Duan, G. *et al.* Strong configurational dependence of elastic properties for a binary model metallic glass. *Appl. Phys. Lett.* **89,** 151901 (2006).
19. Blum, V., Hart, G. L. W., Walorski, M. J. & Zunger, A. Using genetic algorithms to map first-principles results to model Hamiltonians: application to the generalized Ising model for alloys. *Phys. Rev. B* **72,** 165113 (2005).
20. Hart, G. L. W., Blum, V., Walorski, M. J. & Zunger, A. Evolutionary approach for determining first-principles hamiltonians. *Nature Mater.* **4,** 391–394 (2005).
21. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).
22. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50,** 17953–17979 (1994).
23. Methfessel, M. & Paxton, A. T. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B* **40,** 3616–3621 (1989).
24. Blöchl, P. E., Jepsen, O. & Andersen, O. K. Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B* **49,** 16223–16233 (1994).
25. Zhou, Y. *et al.* First-principles studies of Ni-Ta intermetallic compounds. *J. Solid State Chem.* **187,** 211–218 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements Funding by the DFG (Deutsche Forschungsgemeinschaft) grant Mu1648/5 is acknowledged. We also thank the RRZ-Hamburg super-computing site for a generous amount of computational time and E. Kahnert and her team for support and advice related to the computing facilities. Additional computing resources from the German super-computing alliance HLRN are acknowledged.

Author Contributions S.B.M. performed the density functional theory calculations for the NiW, NiTa and CuAl alloys, and M.H. for the NiAl alloy. S.B.M. calculated the cluster expansion Hamiltonians, performed the data post processing and wrote the paper. S.M. formulated the original problem and supervised the investigation. All authors participated in the manuscript preparation during all stages of the process.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M. (stefan.mueller@tuhh.de).

LETTER

doi:10.1038/nature11593

Rapid coupling between ice volume and polar temperature over the past 150,000 years

K. M. Grant[1], E. J. Rohling[1,2], M. Bar-Matthews[3], A. Ayalon[3], M. Medina-Elizalde[1]†, C. Bronk Ramsey[4], C. Satow[5] & A. P. Roberts[2]

Current global warming necessitates a detailed understanding of the relationships between climate and global ice volume. Highly resolved and continuous sea-level records are essential for quantifying ice-volume changes. However, an unbiased study of the timing of past ice-volume changes, relative to polar climate change, has so far been impossible because available sea-level records either were dated by using orbital tuning or ice-core timescales, or were discontinuous in time. Here we present an independent dating of a continuous, high-resolution sea-level record[1,2] in millennial-scale detail throughout the past 150,000 years. We find that the timing of ice-volume fluctuations agrees well with that of variations in Antarctic climate and especially Greenland climate. Amplitudes of ice-volume fluctuations more closely match Antarctic (rather than Greenland) climate changes. Polar climate and ice-volume changes, and their rates of change, are found to covary within centennial response times. Finally, rates of sea-level rise reached at least 1.2 m per century during all major episodes of ice-volume reduction.

During the past few million years, variability in global ice volume (sea level) has been one of the main feedback mechanisms in climate change (see, for example, refs 3, 4). However, detailed assessment of the role of ice volume in climate change is hindered by inadequacies in sea-level records and/or their chronologies. First, dated coral sea-level benchmarks are discontinuous before the last glacial maximum (LGM; ~22,000 years ago). Second, continuous sea-level records have insufficient chronological control; they rely on orbital tuning, correlations with ice-core records, or imperfect transfer of coral datings[1,2,5–7]. Orbital tuning assumes a systematic response between changes in ice volume and Earth's orbital parameters, so that the relationship between insolation forcing and global ice volume cannot be discerned from orbitally tuned records. In addition, the timing of any centennial-scale to millennial-scale fluctuations in ice volume will be poorly constrained in orbitally tuned sea-level records because the shortest orbital frequency is ~19,000 years. Transferring an ice-core chronology to a sea-level record requires an assumption that ice volume always varies in a systematic phase relationship with either Antarctic or Greenland climate, which may not be the case (see, for example, refs 1, 2, 8, 9).

We resolve these issues for the past 150,000 years using a novel approach to provide a detailed chronology to the continuous and highly resolved record of Red Sea relative sea-level (RSL)[2]. We exploit a 'basin isolation' concept, similar to that used for the Red Sea[1,2], in the nearby eastern Mediterranean, where marine sediments can be dated much more accurately. Because the hydrological cycle directly links the $\delta^{18}O$ of eastern Mediterranean surface waters and that of cave speleothems on bordering land masses downwind of this highly evaporative sea[10,11], we can directly relate our new high-resolution planktonic foraminiferal $\delta^{18}O$ record for the surface-dwelling species *Globigerinoides ruber* (white form) in eastern Mediterranean sediment core LC21 ($\delta^{18}O_{ruber}$) to the U–Th-dated Soreq Cave speleothem $\delta^{18}O$ record ($\delta^{18}O_{speleo}$) (Fig. 1; Methods and Supplementary Information).

Previous work demonstrated quantitatively that eastern Mediterranean $\delta^{18}O$ has a strong overprint of sea-level variability[12]; hence, considerable agreement is expected between $\delta^{18}O$ signals for the Red Sea and the eastern Mediterranean (Supplementary Information). Although the more complicated hydrological cycle in the Mediterranean (relative to the Red Sea) means that variations in eastern Mediterranean $\delta^{18}O$ cannot be used to determine the amplitudes of sea-level change precisely, the basin isolation effect imposes sufficient $\delta^{18}O$ signal similarity between the two seas to allow accurate transfer of the superior eastern Mediterranean chronology to the Red Sea record. This is achieved using our new $\delta^{18}O$ record from core LC21 for the subsurface-dwelling planktonic foraminifer *Neogloboquadrina pachyderma* (dextral) ($\delta^{18}O_{pac}$), which is known to minimize surface-water $\delta^{18}O$ overprints (Fig. 1, Methods and Supplementary Information).

Construction of the new RSL chronology involves two stages. First, we build an age model for eastern Mediterranean core LC21 by correlating its $\delta^{18}O_{ruber}$ record with the Soreq Cave $\delta^{18}O_{speleo}$ record over the interval 40–160 kyr BP (Fig. 1a and Supplementary Information). For the interval 0–40 kyr BP, our age model is constrained by 14 radiocarbon datings and two well-documented and independently dated tephra horizons from the Minoan[13] and Campanian Ignimbrite (CI)[14] eruptions. A Bayesian depositional model (constructed using OxCal[15]) comprising the original Soreq Cave chronology, the ^{14}C datings and the chronostratigraphic position of all Soreq–LC21 tie-points (Supplementary Information) then improves the accuracy of the Soreq Cave chronology and ^{14}C datings, and consequently that of core LC21, and rigorously determines the chronological uncertainties of the LC21 tie-points. Next, we transfer the new LC21 age model to the RSL record between 22,000 and 150,000 years ago using the $\delta^{18}O_{pac}$ record (Fig. 1b; Supplementary Information). For the younger interval (0–22,000 years ago), we correlate RSL with a recent sea-level probability curve based on radiometrically dated sea-level benchmarks[16]; this is a more direct correlation target than $\delta^{18}O_{speleo}$ for this interval (Fig. 1b). Our correlations reveal that Last Interglacial (LIG) sea levels peaked before the main (monsoonal) wet phase in the eastern Mediterranean (Fig. 1b). This is stratigraphically corroborated within Red Sea core KL09, in which runoff-related soil biomarkers appear after the LIG highstand signal[17] (Fig. 1b).

Age uncertainties are quantified for all correlations to allow full error propagation into the new RSL chronology. A root-mean-squares estimate at the 95% (2σ) probability level is calculated that fully accounts for errors associated with sample-spacing in the $\delta^{18}O_{speleo}$, $\delta^{18}O_{ruber}$, $\delta^{18}O_{pac}$ and RSL records, as well as the analytical error associated with the Soreq Cave U–Th and LC21 ^{14}C datings, and the 2σ confidence interval of the sea-level probability curve (Supplementary Information). We reinforce this by categorizing our chosen tie-points into three levels of confidence: category 1 is considered the most reliable and within the bounds of sample-spacing, category 2 tie-points may be moved by ±0.5 kyr, and category 3 tie-points are the most

---

[1]School of Ocean and Earth Science, University of Southampton, National Oceanography Centre, European Way, Southampton SO14 3ZH, UK. [2]Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia. [3]Geological Survey of Israel, 30 Malchei Israel Street, Jerusalem 95501, Israel. [4]Research Laboratory for Archaeology and the History of Art, Dyson Perrins Building, University of Oxford, South Parks Road, Oxford OX1 3QY, UK. [5]Department of Geography, Queens Building, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK. †Present address: Centro de Investigación Científica de Yucatán, Unidad Ciencias del Agua, Calle 8, No. 39, Mz. 29, S.M. 64 Cancun, Quintana Roo, CP 77500, México.

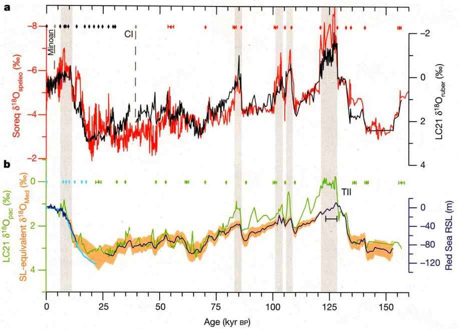


Figure 1 | Correlation of Soreq Cave and eastern Mediterranean (LC21) $\delta^{18}O$ signals and the Red Sea RSL record. a, New planktonic foraminiferal (*G. ruber*) $\delta^{18}O_{ruber}$ record from core LC21 (black), Soreq Cave $\delta^{18}O_{speleo}$ record (red) (Supplementary Information) and tie-points (red diamonds) used to correlate the LC21 and Soreq Cave records. Error bars (±0.5 and ±1 kyr) denote more ambiguous tie-points (Supplementary Information). Also indicated are ^{14}C datings (black diamonds), the Minoan and Campanian Ignimbrite (CI) tephra horizons (dashed black lines) and intervals of sapropel deposition (grey rectangles). **b,** Red Sea RSL record[2] (dark blue, 1-kyr moving Gaussian filter) after correlation with the LC21 (*N. pachyderma*) $\delta^{18}O_{pac}$ record (green) and with a highest-probability sea-level curve[16] (light blue). Correlation tie-points (green and light blue diamonds, with error bars as in **a**) and termination II (TII) are indicated. For a Mediterranean $\delta^{18}O$ ($\delta^{18}O_{Med}$) enrichment of 2.5 ± 0.5‰ per 120 m sea-level change[12], RSL was converted into equivalent $\delta^{18}O_{Med}$ values (orange shading). A LIG wet phase in the Red Sea (about 124–128 kyr BP; black bar) is also indicated[17]. RSL is not reliable for about 14–23 kyr BP because of an aplanktonic zone in Red Sea sediments (Supplementary Information).

contentious and may be moved by ±1 kyr (Fig. 1). On the basis of the total error of each RSL tie-point, we interpolate a 2σ age uncertainty for every data point in the RSL record (Supplementary Information). Finally, these age uncertainties are combined with methodological sea-level uncertainties (±12 m at the 2σ level[1]) in a probabilistic assessment of the RSL record (Fig. 2 and Supplementary Information). This reveals that, during the LIG, RSL at Hanish sill (gateway to the Red Sea) stood above 0 m at 126–130 or 120–133 kyr BP (95% confidence limits to the maximum-probability RSL (RSL_{Pmax}) and RSL data points, respectively), and peaked at 127–129 or 126–132 kyr BP (95% confidence limits to RSL_{Pmax} and RSL data points, respectively; Fig. 2). Although the depth of Hanish sill is implicit in the Red Sea sea-level method, the timing and magnitude of LIG sea levels in our RSL record may be expected to differ from eustatic sea level (ESL) as a result of isostatic effects at the sill[18]; our datings are therefore likely to be refined by detailed isostatic modelling of the sill.

We now compare our probabilistic sea-level curve with other key records of sea level and high-latitude climate (Fig. 2). Our new RSL record agrees well—within uncertainties—with coral sea-level benchmarks (Fig. 2b). Discrepancies in Marine Isotope Stages 5e and 5a may relate to uncertainties in coral position or depth habitat, tectonic/isostatic effects among sites, and/or isostatic effects at Hanish sill[18].

In general, our sea-level record agrees well with variability in ice volume suggested by deep-sea benthic foraminiferal $\delta^{18}O$, and with the major (orbital-scale) climate transitions recorded in Greenland and Antarctic ice cores (Fig. 2). Exceptions to this broad coherence are termination I in the benthic foraminiferal $\delta^{18}O$ record[5] of marine core MD95-2042 from the Iberian margin (Fig. 2d), and termination II and the Marine Isotope Stages 4–3 transition in a global benthic $\delta^{18}O$ stack[19] (Fig. 2c). Given that RSL agrees well with all other proxy records over these transitions, we surmise that these offsets are due to orbital tuning, lower sample resolution[19] and potential bias from deep-sea temperature changes[5] and isostatic effects.

Given that the eustatic glacial–interglacial sea-level range is implicitly accounted for in the Red Sea sea-level method, RSL is a good approximation of variations in ESL (ice volume)[1], although it may underestimate ESL variability by as much as 10% (ref. 18). Regarding polar climate variations, the structure and amplitude of Antarctic climate variations agree well with the record of highest-probability ice-volume fluctuations (Fig. 2e). This corroborates previous observations[1,2,5] but, crucially, is more conclusive because our new chronology is entirely independent of ice-core age models. The Antarctic–RSL relationship is most tenuous at about 95–115 kyr BP, at which RSL instead agrees better with Greenland climate fluctuations. Indeed, the timing of ice-volume changes is generally found to be close to that of Greenland climate variability (Fig. 2f). However, higher-amplitude Greenland $\delta^{18}O$ oscillations ('Dansgaard–Oeschger' events[20]) generally exceed concomitant ice-volume variability. In summary, at the maximum probability and 95% confidence levels for RSL, the timing and structure of large-scale sea-level variability reflects a global signature of climate changes recorded in both Antarctic and Greenland ice cores.

Phase relationships between changes in polar climate and ice volume are initially evaluated by lagged correlations between the ice-core data (European Project for Ice Coring in Antarctica (EPICA) Dronning Maud Land (EDML) $\delta^{18}O$ and North Greenland Ice-core Project (NGRIP) $\delta^{18}O$) and RSL (Supplementary Information). We find that ice-volume lags of 100–400 and 200–400 years produce the best correlations with Antarctic and Greenland climate changes, respectively (Fig. 3a, b and Supplementary Information). Rates of change in ice volume and in polar climate correlate most strongly within ±200 years (Fig. 3a, b). Further assessment with cross-spectral analyses of the EDML $\delta^{18}O$, NGRIP $\delta^{18}O$, and RSL records (and their derivatives) confirms that, for suborbital frequencies, peak coherences between Greenland climate and ice volume are associated with minimal phase offsets (±300 years), whereas phase offsets between Antarctic temperature and ice volume are potentially larger (400–700 years;

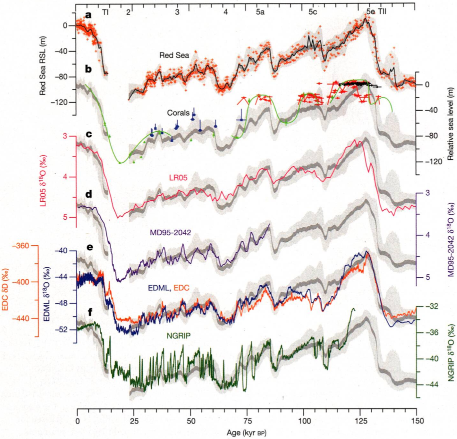


Figure 2 | Comparison of probabilistic assessment of RSL with other sea-level reconstructions and with Antarctic and Greenland climate variability. Confidence intervals of 95% for the RSL data (light grey) and probability maximum (dark grey) (Supplementary Information) are superimposed on: **a**, Red Sea RSL data on our new chronology (orange crosses; black line, 1 kyr moving Gaussian filter); **b**, coral sea-level data ($\pm 2\sigma$) (blue[24], green[25], red[26], black[27]); **c**, a global benthic foraminiferal δ^{18}O stack[19] (pink); **d**, benthic foraminiferal δ^{18}O record (five-point running mean) from marine core MD95-2042 (ref. 5) (purple); **e**, δ^{18}O record (seven-point running mean) from EPICA Dronning Maud Land (EDML)[28] (blue) and δD record (seven-point running mean) from EPICA Dome C (EDC)[29] (orange); and **f**, NGRIP δ^{18}O record (five-point running mean)[30] (green). The MD95-2042 δ^{18}O record is plotted here on the Greenland Ice Core Chronology (GICC05 (ref. 31) for 0–60 kyr BP, and on the NGRIP (2004) chronology[30] for 60–85 kyr BP, after synchronizing the co-registered (MD95-2042) planktonic foraminiferal δ^{18}O record with Greenland δ^{18}O variations. EDML δ^{18}O and EDC δD are plotted on the EDML1/EDC3 timescale[32]. NGRIP δ^{18}O is plotted on the GICC05 timescale for 0–60 kyr BP, and on its original timescale for 60–122 kyr BP (ref. 30). RSL is less reliable for about 14–23 kyr BP because of poor sampling resolution through an aplanktonic interval, and is therefore not shown. Marine Isotope Stages and terminations I and II (TI, TII) are indicated at the top of the figure.

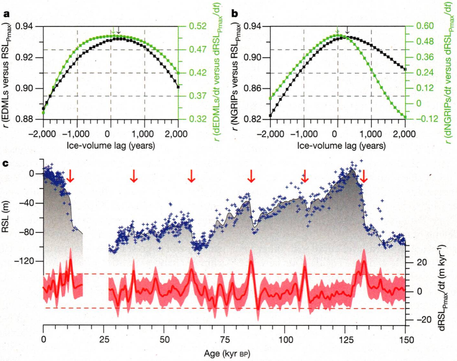


Figure 3 | Lagged correlations of Antarctic and Greenland climate versus ice volume (sea-level), and rates of sea-level change over the last full glacial cycle. **a, b**, Regression coefficients (r) (Supplementary Information) are plotted for the highest-probability sea-level curve and the smoothed (s) EDML and NGRIP δ^{18}O records (RSL$_{\mathrm{Pmax}}$, EDMLs, NGRIPs; black squares, left-hand y axes) and for their first derivatives (dRSL$_{\mathrm{Pmax}}$/dt, dEDMLs/dt, dNGRIPs/dt; green squares; right-hand y axes) for the regressions EDMLs versus RSL$_{\mathrm{Pmax}}$ and dEDMLs/dt versus dRSL$_{\mathrm{Pmax}}$/dt (**a**) and NGRIPs versus RSL$_{\mathrm{Pmax}}$ and dNGRIPs/dt versus dRSL$_{\mathrm{Pmax}}$/dt (**b**). Negative values of ice-volume lag correspond to changes in ice volume leading changes in polar climate. Optimum correlations are indicated (black and green arrows). **c**, RSL$_{\mathrm{Pmax}}$ (grey shading), RSL data (blue crosses) and probability maximum of the first derivative of RSL (red) with 95% confidence interval (pink shading). Rates of sea-level change of +12 and −8 m kyr^{-1} are indicated (dashed lines). Red arrows mark peaks in sea-level rises of more than 12 m kyr^{-1} at 10.9–11.8 kyr BP, 37.4–37.5 kyr BP, 61.2–61.6 kyr BP, 85.5–86.9 kyr BP, 108.1–108.8 kyr BP and 132.1–133.8 kyr BP. Data from the Red Sea aplanktonic interval (about 14–23 kyr BP) are omitted.

Supplementary Information). We infer that Greenland climate closely tracks and/or is directly coupled with ice-volume changes, whereas Antarctic climate variability may lead ice-volume changes by up to 700 years (Supplementary Information).

Our new RSL chronology permits the first robust calculation of rates of relative sea-level change throughout the past 150,000 years (Fig. 3c). This reveals that rates of sea-level rise reached at least 1.2 m per century during all major phases of ice-volume reduction, and were typically up to 0.7 m per century (possibly higher, given the smoothing in our method) when sea-level exceeded 0 m during the LIG (Fig. 3c); the latter is consistent with independent estimates[21,22]. Rates of sea-level lowering rarely exceeded 0.8 m per century. Any differences between rates of change in ESL and RSL at Hanish Sill are likely to be captured within our uncertainties.

We have characterized and dated a continuous record of ice-volume variability throughout the last glacial cycle in a manner that is entirely independent of assumptions about the orbital insolation forcing of climate or about glaciological processes. We have also shown that, on suborbital timescales, polar climate and ice-volume changes were closely coupled in a quasi-direct phase relationship (within centuries). Our analyses hint that Antarctic climate change leads global ice-volume change by several centuries, which is a realistic timescale for ice-sheet adjustments[23]. Greenland climate, however, is found to change virtually simultaneously with ice volume, which may suggest a link of Greenland temperature to ice-volume change in the Northern Hemisphere through albedo feedback.

METHODS SUMMARY

For the Soreq Cave record, we present 440 new U–Th datings that were acquired by multi-collector inductively coupled plasma mass spectrometry at the Geological Survey of Israel (Supplementary Information). Sample ages were corrected for detrital ^{230}Th if ^{230}Th/^{232}Th activity ratios were less than 160; for ^{230}Th/^{232}Th activity ratios of more than 160–200 the correction factor was well within age uncertainties. Typical age uncertainties (2σ) are less than 1 kyr (0–35 kyr BP), less than 1.5 kyr (35–70 kyr BP), less than 2.5 kyr (70–120 kyr BP) and less than 4 kyr (120–160 kyr BP). Dating methods are further described in Supplementary Information. We applied Bayesian age modelling[15] to all U–Th datings ($\pm 2\sigma$), which typically reduced uncertainties to less than 0.5 kyr (0–60 kyr BP), less than 1 kyr (60–90 kyr BP) and less than 2 kyr (90–160 kyr BP) and had only minor impacts on absolute ages (generally less than 250 and less than 750 years for 0–70 and 70–160 kyr BP, respectively).

For the eastern Mediterranean record, continuous u-channel samples from the pristine archive half of core LC21 (35° 40′ N, 26° 35′ E; cruise MD81) were subsampled at 1-cm intervals. Stable isotope analyses of about 15–30 cleaned, similar-sized tests of *G. ruber* (white form) and *N. pachyderma* (dextral) from the greater than 300-μm and 150–300-μm sieved sediment fractions, respectively, were performed at the National Oceanography Centre, Southampton, on a Europa Geo2020 mass spectrometer fitted with an individual acid-bath carbonate preparation line. Standards (NBS-19 and an in-house Carrara marble) were run every 17 samples; external precision is less than 0.06‰.

Received 31 May; accepted 13 September 2012.

Published online 14 November 2012.

1. Siddall, M. *et al.* Sea-level fluctuations during the last glacial cycle. *Nature* **423,** 853–858 (2003).
2. Rohling, E. J. *et al.* Antarctic temperature and global sea level closely coupled over the past five glacial cycles. *Nature Geosci.* **2,** 500–504 (2009).
3. Hansen, J. *et al.* Climate change and trace gases. *Phil. Trans. R. Soc. A* **365,** 1925–1954 (2007).
4. Köhler, P. *et al.* What caused Earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity. *Quat. Sci. Rev.* **29,** 129–145 (2010).
5. Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15,** 565–569 (2000).
6. Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21,** 295–305 (2002).
7. Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437,** 125–128 (2005).
8. Cuffey, K. M. & Marshall, S. J. Substantial contribution to sea-level rise during the last interglacial from the Greenland ice sheet. *Nature* **404,** 591–594 (2000).
9. Otto-Bliesner, B. L. *et al.* Simulating Arctic climate warmth and icefield retreat in the Last Interglacial. *Science* **311,** 1751–1753 (2006).
10. Bar-Matthews, M., Ayalon, A., Gilmour, M., Matthews, A. & Hawkesworth, C. J. Sea–land oxygen isotopic relationships from planktonic foraminifera and speleothems in the Eastern Mediterranean region and their implication for paleorainfall during interglacial intervals. *Geochim. Cosmochim. Acta* **67,** 3181–3199 (2003).
11. Almogi-Labin, A. *et al.* Climatic variability during the last ~90 ka of the southern and northern Levantine Basin as evident from marine records and speleothems. *Quat. Sci. Rev.* **28,** 2882–2896 (2009).
12. Rohling, E. J. Environmental controls on salinity and δ^{18}O in the Mediterranean. *Paleoceanography* **14,** 706–715 (1999).
13. Manning, S. W. *et al.* Chronology for the Aegean Late Bronze Age 1700–1400 BC. *Science* **312,** 565–569 (2006).
14. De Vivo, B. *et al.* New constraints on the pyroclastic eruptive history of the Campanian volcanic plain (Italy). *Mineral. Petrol.* **73,** 47–65 (2001).
15. Bronk Ramsey, C. Deposition models for chronological records. *Quat. Sci. Rev.* **27,** 42–60 (2008).
16. Stanford, J. D. *et al.* Sea-level probability for the last deglaciation: a statistical analysis of far-field records. *Global Planet. Change* **79,** 193–203 (2011).
17. Trommer, G. *et al.* Sensitivity of Red Sea circulation to sea level and insolation forcing during the last interglacial. *Clim. Past* **7,** 941–955 (2011).
18. Lambeck, K. *et al.* Sea level and shoreline reconstructions for the Red Sea: isostatic and tectonic considerations and implications for hominin migration out of Africa. *Quat. Sci. Rev.* **30,** 3542–3574 (2011).
19. Lisiecki, L. E. & Raymo, M. E. A Plio-Pleistocene stack of 57 globally distributed benthic δ^{18}O records. *Paleoceanography* **20,** PA1003, doi:10.1029/2004PA001071 (2005).
20. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364,** 218–220 (1993).
21. Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C. & Oppenheimer, M. Probabilistic assessment of sea level during the last interglacial stage. *Nature* **462,** 863–868 (2009).
22. Thompson, W. G., Curran, H. A., Wilson, M. A. & White, B. Sea-level oscillations during the last interglacial highstand recorded by Bahamas corals. *Nature Geosci.* **4,** 684–687 (2011).
23. Gregory, J. M., Huybrechts, P. & Raper, S. C. B. Threatened loss of the Greenland ice-sheet. *Nature* **428,** 616 (2004).
24. Chappell, J. Sea level changes forced ice breakouts in the Last Glacial cycle: new results from coral terraces. *Quat. Sci. Rev.* **21,** 1229–1240 (2002).
25. Cutler, K. B. *et al.* Rapid sea-level fall and deep-ocean temperature change since the last interglacial period. *Earth Planet. Sci. Lett.* **206,** 253–271 (2003).
26. Thompson, W. G. & Goldstein, S. L. Open-system coral ages reveal persistent suborbital sea-level cycles. *Science* **308,** 401–404 (2005).
27. Dutton, A. & Lambeck, K. Ice volume and sea level during the last interglacial. *Science* **337,** 216–219 (2012).
28. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444,** 195–198 (2006).
29. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317,** 793–797 (2007).
30. North Greenland Ice Core Project Members. High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431,** 147–151 (2004).
31. Svensson, A. *et al.* A 60 000 year Greenland stratigraphic ice core chronology. *Clim. Past* **4,** 47–57 (2008).
32. Parrenin, F. *et al.* The EDC3 chronology for the EPICA Dome C ice core. *Clim. Past* **3,** 485–497 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Dutton for comments that improved the manuscript. S. Lee helped with the use of OxCal. This study contributes to UK Natural Environment Research Council (NERC) projects NE/H004424/1, NE/E01531X/1 and NE/I009906/1, to a Royal Society Wolfson Research Merit Award (E.J.R.), and to a 2012 Australian Laureate Fellowship FL120100050 (E.J.R.).

Author Contributions K.M.G. led the study. E.J.R. designed the study, contributed to statistical analyses and co-wrote the paper. M.B.-M. and A.A. developed the Soreq Cave speleothem record. M.M.E. contributed to the interpretations. C.B.R. supported the Bayesian age modelling. C.S. contributed the LC21 tephrachronology. A.P.R. contributed to the discussion of results and manuscript refinement.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M.G. (kxg@noc.soton.ac.uk).

LETTER

doi:10.1038/nature11555

Development of teeth and jaws in the earliest jawed vertebrates

Martin Rücklin[1], Philip C. J. Donoghue[1], Zerina Johanson[2], Kate Trinajstic[3,4], Federica Marone[5] & Marco Stampanoni[5,6]

Teeth and jaws constitute a model of the evolutionary developmental biology concept of modularity[1] and they have been considered the key innovations underpinning a classic example of adaptive radiation[2]. However, their evolutionary origins are much debated. Placoderms comprise an extinct sister clade[3] or grade[4,5] to the clade containing chondrichthyans and osteichthyans, and although they clearly possess jaws, previous studies have suggested that they lack teeth[6–8], that they possess convergently evolved tooth-like structures[9–11] or that they possess true teeth[12]. Here we use synchrotron radiation X-ray tomographic microscopy (SRXTM)[13] of a developmental series of *Compagopiscis croucheri* (Arthrodira) to show that placoderm jaws are composed of distinct cartilages and gnathal ossifications in both jaws, and a dermal element in the lower jaw. The gnathal ossification is a composite of distinct teeth that developed in succession, polarized along three distinct vectors, comparable to tooth families. The teeth are composed of dentine and bone, and show a distinct pulp cavity that is infilled centripetally as development proceeds. This pattern is repeated in other placoderms, but differs from the structure and development of tooth-like structures in the postbranchial lamina and dermal skeleton of *Compagopiscis* and other placoderms. We interpret this evidence to indicate that *Compagopiscis* and other arthrodires possessed teeth, but that tooth and jaw development was not developmentally or structurally integrated in placoderms. Teeth did not evolve convergently among the extant and extinct classes of early jawed vertebrates but, rather, successional teeth evolved within the gnathostome stem-lineage soon after the origin of jaws. The chimaeric developmental origin of this model of modularity reflects the distinct evolutionary origins of teeth and of component elements of the jaws.

Possible scenarios for the evolutionary origin of teeth and jaws have been influenced heavily by chondrichthyans, in which teeth develop within a deep dental lamina, producing files of replacements preformed long ahead of their functional deployment[6]. However, chondrichthyans are not primitive jawed vertebrates[4] but, along with osteichthyans, represent crown-group gnathostomes. Therefore, the pattern of tooth development and replacement in living chondrichthyans does not necessarily reflect the nature of the earliest jawed vertebrates. The extinct placoderms are the most primitive jawed vertebrates known, comprising either a monophyletic sister lineage to crown gnathostomes[3] (Fig. 1a) or, more persuasively, a primitive grade of jawed vertebrates that includes a succession of sister lineages to crown gnathostomes[4,5] (Fig. 1b). As such, placoderms are crucial to resolving the early evolution of teeth and jaws.

The nature of the dentition in placoderms has been the subject of debate that can be reduced to semantic differences, with some authors advocating a structural diagnosis that identifies teeth in placoderms[11,12] and others adhering to the use of developmental criteria that preclude the presence of teeth in placoderms[6–8]. Intuitively, developmental definitions cannot be applied to fossil material, but the pattern of skeletal development is preserved in the sclerochronology of growth-arrest lines and the polarity of cell lacunae and canaliculi in mineralised skeletal tissues. So far, however, understanding of placoderm jaw and dental development has been limited largely to inference from surface morphology[7], with only one direct study of development[12]. This is because traditionally analyses have been destructive. We used SRXTM[13] to study jaw, dental and dermal skeletal development in ontogenetic stages of the arthrodire *Compagopiscis croucheri*, selected because it is known from abundant articulated three-dimensional remains; the smallest and largest specimens known were included in our study.

The lower jaw of *Compagopiscis* is comprised of the infragnathal that rests on the Meckel's cartilage which is ossified only at its proximal (articular) and distal (mentomeckelian) extremes (Supplementary Figs 1 and 2). Tomographic data reveal that the infragnathal is composed of two principal ossifications, the bony shaft of the jaw (axial ossification) and a distal compound dental ossification (Fig. 2). We digitally extracted successive stages of growth in the infragnathal, revealing the sequence of development of the dental ossification (Figs 2d, e, g, h and 3). Growth proceeded through the addition of cusps, proximally, distally and lingually relative to a primordial cusp. Initial addition consists of single cusps and subsequent cusps are associated with the growth of an increasingly expansive sheet of tissue that in later developmental stages extends ventrally around the bony shaft of the infragnathal and, in the largest infragnathal, partially around the Meckel's cartilage (Figs 2 and 3). In most cases, these sheets of tissue are continuous from the proximal to distal rows of cusps (Fig. 3), indicating coordinated growth. More cusps are added to the proximal row than to the distal and lingual rows during each growth episode. However, the

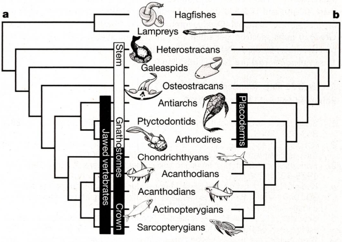


Figure 1 | Evolutionary relationships of principal groups of vertebrates. **a**, **b**, Phylogenetic relationships among the principal groups of stem and crown gnathostomes. The traditional view of placoderm monophyly[27,28] (**a**) versus the more recent hypothesis of placoderm paraphyly[4] (**b**).

---

[1]School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK. [2]Earth Sciences, The Natural History Museum, Cromwell Road, South Kensington, London SW7 5BD, UK. [3]Department of Chemistry, Curtin University, Kent St, Bentley 6102, Australia. [4]Department of Earth and Planetary Sciences, Western Australian Museum, Kew Street, Welshpool 6106, Australia. [5]Swiss Light Source, Paul Scherrer Institut, CH-5232 Villigen, Switzerland. [6]Institute for Biomedical Engineering, University and ETH Zürich, CH-8092 Zürich, Switzerland.

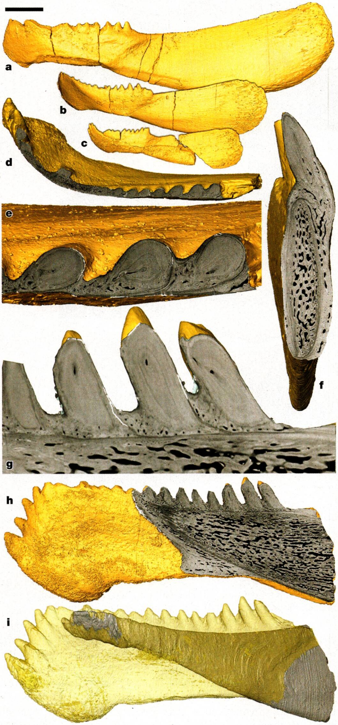


Figure 2 | Lower-jaw element of *Compagopiscis croucheri*, Late Devonian period, Australia. **a–c**, microCT data; **d–i**, SRXTM data. Volume rendering of jaws and teeth (**a–c**, **i**) and surface cut (**d–h**). Ontogenetic sequence in lateral view (**a**) NHMUK PV P.50948, (**b**) NHMUK PV P.50943 and WAM 91.4.3 (**c**). Teeth and jaw ossifications WAM 91.4.3: horizontal section (**d**, **e**), vertical section (**f**), longitudinal section (**g**, **h**) and labelled sclerochronology as virtual dental ossification (transparent) and bony shaft (shaded, **i**). Scale bar in **a** represents 2 mm (**a–c**), 1 mm (**d**, **h**, **i**), 240 μm (**f**), 400 μm (**e**, **g**).

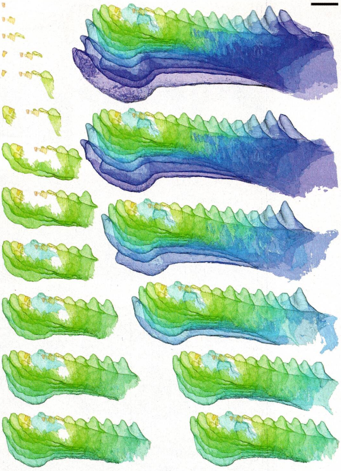


Figure 3 | Virtual development of a *Compagopiscis croucheri* lower jaw, Late Devonian period, Australia. Labelled sclerochronology in the dental element of the lower jaw of WAM 91.4.3. Subsequent ontogenetic stages follow from top left to top right in an anti-clockwise direction. Darker shades represent the addition of developmental stages of teeth and sheets of tissue. Scale bar, 1 mm.

individual cusps are clearly successional (not synchronous) even within each growth episode and tooth row (Figs 2d–g and 3). The cusps are composed of dentine; they have distinct pulp cavities lined with centripetally nested tissue layers permeated by radially arranged and polarized canaliculi, but not cell lacunae (Figs 2e, g and 4a). The pulp cavity of each cusp is infilled progressively through ontogeny such that the earliest formed cusps are completely infilled. The structure and pattern of development of the supragnathals (Fig. 4a), which attach directly to the roof of the oral cavity, are comparable to the dental ossification of the infragnathal.

The pattern of development of the dental ossification in *Compagopiscis* is compatible with observations made in other arthrodiran placoderms. For example, we can corroborate the identification of distinct dental and axial ossifications comprising the infragnathal of *Plourdosteus*, as well as the coordination of proximal and distal cusp development that was inferred but could not be observed in the same study[12]. However, we find no evidence for the widespread resorption and remodelling in *Compagopiscis* that was inferred for *Plourdosteus*[12]. An axial and dental ossification, along with sequential cusp development in a proximal row, is present in the infragnathal of buchanosteids (Fig. 4b), which are distantly related arthrodires[14], indicating that these characteristics are primitive features of arthrodires. Distinct gnathal and dermal jaw ossifications seem to be primitive for placoderms more generally, as a cusp-bearing capping structure occurs associated with a proximal axial ossification in rhenanids[15]. An equivalent of the dermal axial ossification in arthrodires occurs also in antiarchs (Fig. 4c), in which the secondary absence of dentine cusps from the infragnathal mirrors their secondary absence from the dermal skeleton[16]. However, it may alternatively reflect a primitive absence of teeth deep within the paraphyletic placoderm grade of stem-gnathostomes[4,5].

The structure of the individual cusps and their pattern of sequential development within the gnathal elements in *Compagopiscis* and other

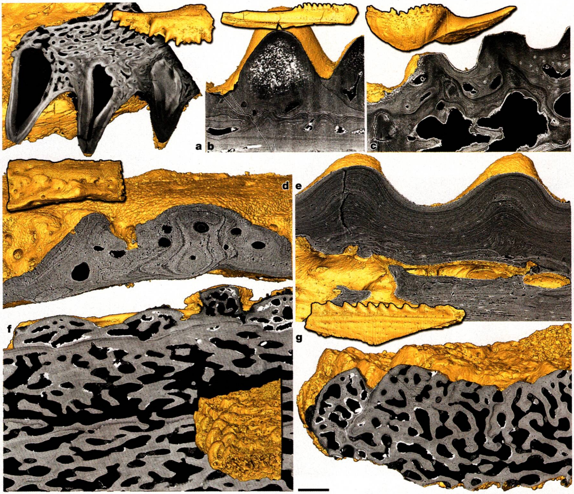


Figure 4 | Histological comparison of the *Compagopiscis croucheri* cusps on jaws, dermal bone and postbranchial lamina, with the jaws and pectoral fins of the antiarch *Bothriolepis* species, both Late Devonian, Australia, and the jaw of a buchanosteid arthrodire, Early Devonian, Saudi Arabia. **a–g**, Volume rendering of microtomography data (small insets; **a**, **b**), volume rendering of SRXTM data (small insets; **c–f**), surface-cuts showing SRXTM images (large insets; **d**, **e**, **g**) and virtual thin sections of SRXTM images (**a–c**, **f**). Cusps on jaws of *Compagopiscis croucheri*, proximal upper jaw (posterior supragnathal) National History Museum London (NHMUK) PV P.57629, small inset NHMUK PV P.50943 (**a**), buchanosteid arthrodire, lower jaw, Muséum National d'Histoire Naturelle Paris MNHN.F.ARB 239 (**b**) and *Bothriolepis* sp. lower jaw NHMUK PV P.50898 (**c**). Tubercles on dermal plate (marginal) of *Compagopiscis croucheri* NHMUK PV P.50945 (**d**), marginal cusps on pectoral fin of *Bothriolepis* sp. NHMUK PV P.50898 (**e**) and tooth-like cusps on postbranchial lamina of *Compagopiscis croucheri* NHMUK PV P.5255.6 (**f**, **g**). Scale bar in **g** represents 500 μm (**a**, small inset **f**), 200 μm (**g**, small inset **a**), 285 μm (**b**), 5 mm (small inset **b**), 167 μm (**c–e**), 2 mm (small inset **c**), 400 μm (small inset **d**), 333 μm (small inset **e**), 143 μm (**f**).

placoderms, support comparisons to gnathostome teeth that have been made previously based on surface morphology[9–11]. This interpretation has been contested on the grounds that tooth-like cusps are not diagnostic of teeth, evidenced by the fact that even in placoderms such as *Compagopsicis*, comparable tooth-like cusps also occur in association with the cranial dermal skeleton[3,7,8]. To test this comparison we examined the structure and development of the dermal skeleton. The *Compagopiscis* dermal skeleton is similar to that of other placoderms that have been investigated[16], in that they are composed of a basal division of lamellar bone, a middle division of cancellar bone, and in the superficial division of compact bone with surface tubercles. However, unlike the morphogenetically distinct cusps associated with the gnathals, the superficial tubercles of the dermal skeleton are revealed to be focal developments of continuous sheets of bone that are morphogenetically integrated with the underlying dermoskeleton (Fig. 4d). Tooth-like structures associated with the dermal pectoral fin spines of antiarch placoderms could be an exception; when isolated, these structures can be mistaken for complete jaws[8] (Fig. 4e). However, our analysis reveals that the tooth-like margin of the pectoral spines is again characteristic of the dermoskeleton, comprised collectively of continuous sheets of bone, not from morphogenetically distinct elements as in the gnathals (Figs 2 and 4a, b). Evidently, structural objections to the identification of teeth in placoderms are unfounded.

It has been suggested that the placoderm dentition fails to meet the definition of a gnathostome tooth because there is no evidence that they develop from a deep and continuous dental lamina[3,6,8]. However, living jawed vertebrates show great diversity in dental development, with teeth developing in deep or shallow positions, from continuous to discrete epithelial pockets, that persist through life or atrophy and develop anew[17–19]; the plesiomorphic conditions for jawed vertebrates and crown gnathostomes are unclear. The discrete teeth in *Compagopiscis* and other placoderms developed in a shallow position like those of many living osteichthyans. The key distinction in placoderms is that the dentition is statodont, that is, teeth are not resorbed, shed and subsequently replaced. In this sense, the placoderm dentition is most similar to that of ischnacanthid acanthodians[20], holocephalans[21] and lungfish[22], in which teeth develop through marginal apposition to a compound dental plate. The functional limitation of this approach to development is that worn teeth cannot be replaced. The innovation of crown gnathostomes is site-specific tooth replacement.

It has been suggested that successional tooth homologues are present also on the postbranchial wall of placoderms, including

Compagopiscis[9,10], coopted convergently to the jaw among different lineages of primitive gnathostomes to serve a tooth function[9,10]. The cusps are arranged into columns and rows reminiscent of the tooth families of extant chondrichthyans[9,10], but the hypothesis of sequential development is an inference based on surface morphology. Our data refute this hypothesis, given that in the dermoskeleton, the rows of tooth-like cusps that occur on the postbranchial wall are simple focal developments of continuous sheets of spongy bone, added episodically to the growing margin of the postbranchial wall (Fig. 4f, g).

Our evidence indicates that teeth are present even in the earliest jawed vertebrates and that within the phylogenetic context of placoderm paraphyly[4,5] they can be identified as homologous to the teeth of crown gnathostomes. This contrasts with the hypothesis that teeth were absent from the earliest jawed vertebrates, evolved convergently through cooption of oral and pharyngeal denticles[9–11]. Indeed, our tomographic analyses show that the putative tooth-like pattern of placoderm pharyngeal denticle replacement bears no resemblance to that of their teeth, except in superficial morphology. Thus, the hypothesis of a distinct evolutionary origin of teeth and dermal denticles[9,10,23] can be rejected, as jawless stem-gnathostomes have been shown to lack homologous dental patterning[24] and the assertion of a fundamental embryological distinction between external and oral denticles has been refuted[25]. Ultimately, teeth and other oral and pharyngeal denticles must be derived through the extension of the odontogenic capacity of the external dermis to the internal dermis and endoderm. However, tooth- and jaw-structure and development is evidently less integrated within the placoderm grade of early jawed-vertebrate evolution than in derived osteichthyans in which teeth, tooth development and jaw structure are intimately interwoven, as part of the process of site-specific tooth replacement. Upper and lower dental ossifications occur in placoderms, but there is no clear homologue of the osteichthyan dentary or coronoid. However, the axial ossification of the infragnathal can be compared to the inner dental arcade of early osteichthyans based on its location relative to the underlying Meckel's cartilage, overlying dental ossification and lateral muscle attachment. *Compagopiscis* and other placoderms evidence an early stage in jawed vertebrate evolution in which the components of the mandible were fewer and more obviously distinct than they are in osteichthyan evolutionary and developmental model organisms. Some processes associated with these more derived taxa, such as tooth resorption (as a necessary precursor to tooth replacement), are absent in placoderms. This stepwise acquisition reflects the fact that character complexes like the gnathostome jaw have been assembled over a protracted episode of evolutionary history and so the modular construction of the mandible[1], for example, reflects the disparate evolutionary origins of its component modules.

METHODS SUMMARY

Museum repositories: National History Museum London (NHMUK); Western Australian Museum (WAM); Muséum National d'Histoire Naturelle Paris (MNHN). The material was scanned using a SkyScan 1172 Micro-CT scanner at the University of Bristol, and using the TOMCAT beamline[26] of the Swiss Light Source (SLS), Paul Scherrer Institut, Switzerland. Slice data were analysed and manipulated using Avizo 6.3 (http://www.vsg3d.com).

Full Methods and any associated references are available in the online version of the paper.

Received 3 July; accepted 10 September 2012.

Published online 17 October 2012.

1. Atchley, W. R. & Hall, B. K. A model for development and evolution of complex morphological structures and its application to the mammalian mandible. *Biol. Rev. Camb. Philos. Soc.* **66,** 101–157 (1991).
2. Gans, C. & Northcutt, R. G. Neural crest and the origin of the vertebrates: a new head. *Science* **220,** 268–273 (1983).
3. Young, G. C. Placoderms (armored fish): dominant vertebrates of the Devonian Period. *Annu. Rev. Earth Planet. Sci.* **38,** 523–550 (2010).
4. Brazeau, M. D. The braincase and jaws of a Devonian 'acanthodian' and modern gnathostome origins. *Nature* **457,** 305–308 (2009).
5. Davis, S. P. *et al.* Acanthodes and shark-like conditions in the last common ancestor of modern gnathostomes. *Nature* **486,** 247–250 (2012).
6. Reif, W.-E. Evolution of dermal skeleton and dentition in vertebrates: The odontode regulation theory. *Evol. Biol.* **15,** 287–368 (1982).
7. Burrow, C. J. Comment on "Separate evolutionary origins of teeth from evidence in fossil jawed vertebrates". *Science* **300,** 1661 (2003).
8. Young, G. C. Did placoderm fish have teeth? *J. Vertebr. Paleontol.* **23,** 987–990 (2003).
9. Johanson, Z. & Smith, M. M. Placoderm fishes, pharyngeal denticles, and the vertebrate dentition. *J. Morphol.* **257,** 289–307 (2003).
10. Johanson, Z. & Smith, M. M. Origin and evolution of gnathostome dentitions: a question of teeth and pharyngeal denticles in placoderms. *Biol. Rev. Camb. Philos. Soc.* **80,** 303–345 (2005).
11. Smith, M. M. & Johanson, Z. Separate evolutionary origin of teeth from evidence in fossil jawed vertebrates. *Science* **299,** 1235–1236 (2003).
12. Ørvig, T. Histologic studies of ostracoderms, placoderms and fossil elasmobranchs 3. Structure and growth of gnathalia of certain arthrodires. *Zool. Scr.* **9,** 141–159 (1980).
13. Donoghue, P. C. J. *et al.* Synchrotron X-ray tomographic microscopy of fossil embryos. *Nature* **442,** 680–683 (2006).
14. Lelièvre, H. Description of *Maideria falipoui* n.g., n.sp., a long snouted brachythoracid (Vertebrata, Placodermi, Arthrodira) from the Givetian of Maider (South Morocco), with a phylogenetic analysis of primitive brachythoracids. *Bulletin du Muséum National d'Histoire Naturelle, Paris* **17,** 163–207 (1995).
15. Young, G. C. The relationships of placoderm fishes. *Zool. J. Linn. Soc.* **88,** 1–57 (1986).
16. Downs, J. P. & Donoghue, P. C. J. Skeletal histology of *Bothriolepis canadensis* (Placodermi, Antiarchi) and evolution of the skeleton at the origin of jawed vertebrates. *J. Morphol.* **270,** 1364–1380 (2009).
17. Huysseune, A. *et al.* Evolutionary and developmental origins of the vertebrate dentition. *J. Anat.* **214,** 465–476 (2009) Medline.
18. Huysseune, A. *et al.* Unique and shared gene expression patterns in Atlantic salmon (Salmo salar) tooth development. *Dev. Genes Evol.* **218,** 427–437 (2008).
19. Fraser, G. J. *et al.* Developmental and evolutionary origins of the vertebrate dentition: Molecular controls for spatio-temporal organisation of tooth sites in osteichthyans. *J. Exp. Zool. B Mol. Dev. Evol.* **306,** 183–203 (2006).
20. Ørvig, T. Acanthodian dentition and its bearing on the relationships of the group. *Palaeontographica (Abt. A)* **143,** 119–150 (1973).
21. Finarelli, J. A. & Coates, M. I. First tooth-set outside the jaws in a vertebrate. *Proc. R. Soc. B* **279,** 775–779 (2012).
22. Smith, M. M. The pattern of histogenesis and growth of tooth plates in larval stages of extant lungfish. *J. Anat.* **140,** 627–643 (1985).
23. Fraser, G. J. *et al.* The odontode explosion: the origin of tooth-like structures in vertebrates. *Bioessays* **32,** 808–817 (2010).
24. Rücklin, M. *et al.* Teeth before jaws? Comparative analysis of the structure and development of the external and internal scales in the extinct jawless vertebrate *Loganellia scotica*. *Evol. Dev.* **13,** 523–532 (2011).
25. Soukup, V. *et al.* Dual epithelial origin of vertebrate oral teeth. *Nature* **455,** 795–798 (2008).
26. Stampanoni, M. *et al.* TOMCAT: A beamline for TOmographic Microscopy and Coherent rAdiology experimenTs. *Synchrotron Radiation Instrumentation, Pts 1 and 2* **879,** 848–851 http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=21052651 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Bengtson, J. Cunningham, D. Murdock, S. Giles and A. Hetherington for help at the TOMCAT beamline; K. Robson-Brown for help at the Micro-CT, and H. Lelièvre and G. Clément for loan of specimens. The study was funded by EU grant FP7 MC-IEF (to M.R. and P.C.J.D.), Australian Research Council Grant DP 110101127 (to Z.J. and K.T.), Natural Environment Research Council grant NE/G016623/1 (to P.C.J.D.) and the Paul Scherrer Institut (P.C.J.D.).

Author Contributions M.R., P.C.J.D. and Z.J. conceived the project. K.T. acid-prepared WAM specimens. M.R., P.C.J.D., F.M. and M.S. collected the data. M.R. analysed the data. All authors contributed to the interpretation of the data and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R. (m.ruecklin@bristol.ac.uk) and P.C.J.D. (phil.donoghue@bristol.ac.uk).

doi:10.1038/nature11688

Global convergence in the vulnerability of forests to drought

Brendan Choat[1]*, Steven Jansen[2]*, Tim J. Brodribb[3], Hervé Cochard[4,5], Sylvain Delzon[6], Radika Bhaskar[7], Sandra J. Bucci[8], Taylor S. Feild[9], Sean M. Gleason[10], Uwe G. Hacke[11], Anna L. Jacobsen[12], Frederic Lens[13], Hafiz Maherali[14], Jordi Martínez-Vilalta[15,16], Stefan Mayr[17], Maurizio Mencuccini[18,19], Patrick J. Mitchell[20], Andrea Nardini[21], Jarmila Pittermann[22], R. Brandon Pratt[12], John S. Sperry[23], Mark Westoby[10], Ian J. Wright[10] & Amy E. Zanne[24,25]

Shifts in rainfall patterns and increasing temperatures associated with climate change are likely to cause widespread forest decline in regions where droughts are predicted to increase in duration and severity[1]. One primary cause of productivity loss and plant mortality during drought is hydraulic failure[2–4]. Drought stress creates trapped gas emboli in the water transport system, which reduces the ability of plants to supply water to leaves for photosynthetic gas exchange and can ultimately result in desiccation and mortality. At present we lack a clear picture of how thresholds to hydraulic failure vary across a broad range of species and environments, despite many individual experiments. Here we draw together published and unpublished data on the vulnerability of the transport system to drought-induced embolism for a large number of woody species, with a view to examining the likely consequences of climate change for forest biomes. We show that 70% of 226 forest species from 81 sites worldwide operate with narrow ($<$1 megapascal) hydraulic safety margins against injurious levels of drought stress and therefore potentially face long-term reductions in productivity and survival if temperature and aridity increase as predicted for many regions across the globe[5,6]. Safety margins are largely independent of mean annual precipitation, showing that there is global convergence in the vulnerability of forests to drought, with all forest biomes equally vulnerable to hydraulic failure regardless of their current rainfall environment. These findings provide insight into why drought-induced forest decline is occurring not only in arid regions but also in wet forests not normally considered at drought risk[7,8].

Sensitivity to drought is fundamentally important in shaping the geographic distribution of individual species as well as communities[9,10]. Drought has underpinned many large-scale forest mortality events over the past century, often in combination with other abiotic and biotic factors[1,11]. Recent evidence suggests rising global temperatures are already amplifying drought-induced forest change and affecting terrestrial net primary productivity[12–14]. The consequences of longer droughts and higher temperatures are potentially dramatic. For example, rapid forest collapse as a result of drought could convert the world's tropical forests from a net carbon sink into a large carbon source during this century[8,15]. Predicting how forests will respond to future climate changes hinges on a quantitative understanding of the physiological mechanisms governing drought stress at the species level. One of the most promising avenues for characterizing the sensitivity of plants to drought stress is by quantifying the strength of the liquid (hydraulic) connection between soil and leaves through the water-transporting xylem tissue.

Cavitation, a phase change from liquid water to vapour, occurs in plants because water transported through the xylem is under negative pressure[16]. The resultant air emboli block xylem conduits and reduce the plant's ability to move water from soil to sites of photosynthesis[17]. Recent evidence indicates that the ability of woody plants to survive and recover from periods of sustained drought is strongly related to their embolism resistance[2,3]. This property varies widely among species and is largely determined by differences in the structure of the xylem[18–20]. Although xylem structure can acclimate to environmental variation during growth and development, subsequent acclimation of embolism resistance to environmental stress is not possible because xylem conduits are dead at maturity. Embolism resistance therefore represents a critically important trait for defining the limits of drought tolerance across woody species and predicting drought-induced forest decline at regional and global scales.

In drying soil, stomata initially regulate water loss from the leaves to maintain xylem pressure (Ψ_x; measured as water potential below 0) within a range that will protect the xylem from extensive embolism[17,21]. As drought continues, stomatal closure slows but does not halt the decline of xylem pressure and hydraulic capacity. If soil water is not replenished before complete hydraulic failure occurs then the plant will desiccate and potentially die. The resistance of a plant to embolism is described by the relationship between xylem pressure and loss of hydraulic conductivity due to conduit occlusion by gas emboli (Supplementary Fig. 1). The Ψ_x at which 50% loss of conductivity occurs (Ψ_{50}) is the most commonly used index of embolism resistance. When Ψ_x falls below Ψ_{50} the water transport function of the xylem is markedly impaired and the plant is exposed to considerable risk of accelerated embolism leading to long-term reductions in productivity, tissue damage, and ultimately death[2]. To examine vulnerability of forest biomes to drought-induced hydraulic failure we assembled a database of Ψ_{50} (Supplementary Table 1, including 480 woody species). Site

---

[1]University of Western Sydney, Hawkesbury Institute for the Environment, Richmond, New South Wales 2753, Australia. [2]Ulm University, Institute for Systematic Botany and Ecology, Albert-Einstein-Allee 11, 89081 Ulm, Germany. [3]University of Tasmania, School of Plant Science, Private Bag 55, Hobart, Tasmania 7001, Australia. [4]INRA, UMR547 PIAF, F-63100 Clermont-Ferrand, France. [5]Clermont Université, Université Blaise Pascal, UMR547 PIAF, F-63000 Clermont-Ferrand, France. [6]INRA, University of Bordeaux, UMR BIOGECO, 33450 Talence, France. [7]Brown University, Environmental Change Initiative, Box 1951, 167 Thayer Street, Providence, Rhode Island 02912, USA. [8]Universidad Nacional de la Patagonia San Juan Bosco, Departmento de Biología, Facultad de Ciencias Naturales, 9000 Comodoro Rivadavia, Argentina. [9]James Cook University, School of Marine and Tropical Biology, Townsville, Queensland 4811, Australia. [10]Macquarie University, Department of Biological Sciences, New South Wales 2109, Australia. [11]University of Alberta, Department of Renewable Resources, Edmonton, Alberta T6G 2E3, Canada. [12]California State University, Department of Biology, Bakersfield, California 93311, USA. [13]Naturalis Biodiversity Centre, Leiden University, PO Box 9514, 2300 RA Leiden, The Netherlands. [14]University of Guelph, Department of Integrative Biology, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada. [15]CREAF, Cerdanyola del Vallès 08193, Spain. [16]Universitat Autònoma Barcelona, Cerdanyola del Vallès 08193, Spain. [17]University Innsbruck, Institut für Botank, Sternwartestrasse 15, A-6020 Innsbruck, Austria. [18]ICREA at CREAF, Univ Autònoma Barcelona, Cerdanyola del Vallès 08193, Spain. [19]University of Edinburgh, School of GeoSciences, Crew Building, West Mains Road, Edinburgh EH9 3JN, UK. [20]CSIRO, Ecosystem Sciences, College Road, Sandy Bay, Tasmania 7005, Australia. [21]Università di Trieste, Dipartimento di Scienze della Vita, Via L. Giorgieri 10, 34127 Trieste, Italy. [22]University of California, Santa Cruz, Department of Ecology and Evolutionary Biology, California 95064, USA. [23]University of Utah, Department of Biology, 257 South 1400 East, Salt Lake City, Utah 84112, USA. [24]Missouri Botanical Garden, Center for Conservation and Sustainable Development, St. Louis, Missouri 63166, USA. [25]George Washington University, Department of Biological Sciences, 2023 G Street NW, Washington DC 20052, USA.
*These authors contributed equally to this work.

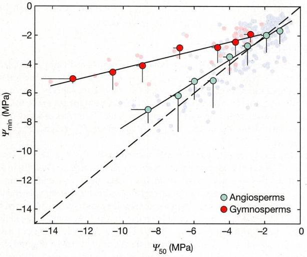


Figure 1 | Minimum xylem pressure as a function of embolism resistance for 191 angiosperm and 32 gymnosperm species. The safety margin is the distance between each point and the 1:1 (dashed) line. Data were binned in 1.0-MPa increments for embolism resistance (Ψ_{50}). Bins were pooled with the next lowest bin if they contained only one sample. Raw data are shown as smaller points behind binned data. Error bars, s.d. Regression lines shown were fitted to raw data (angiosperms, $r^2 = 0.57$, $P < 0.0001$; gymnosperms, $r^2 = 0.59$, $P < 0.0001$; data set available in Supplementary Table 1). Ψ_{min}, minimum xylem pressure.

climate varied widely, for example, mean annual precipitation (MAP) ranged from 300 to 4,500 mm, and mean annual temperature from −4 to 27 °C. Data for angiosperms (flowering plants) and gymnosperms (mainly conifers) were analysed separately because of fundamental differences in xylem structure between these two groups.

We observed a significant ($P < 0.0001$) linear relationship between the minimum Ψ_x measured in plants under natural conditions (Ψ_{min}) and Ψ_{50} for both angiosperms and gymnosperms, showing that embolism resistance is tightly linked with the level of drought stress experienced by plants across a broad range of environments (Fig. 1). The difference between Ψ_{min} and Ψ_{50} represents a highly informative 'safety margin' within which the plant operates in a given environment[22,23]. This safety margin quantifies the degree of conservatism in a plant's hydraulic strategy, indicating that plants with low (or even negative) safety margins experience large amounts of embolism and therefore potential risk of hydraulic failure (Supplementary Fig. 1). Measurements of Ψ_{min} were available for 226 of the 480 species included in our database.

Across all forest biomes, 70% of all species operated at narrow (<1 megapascal (MPa)) safety margins (Fig. 2a), indicating that both arid and mesic biomes are vulnerable to drought-induced decline if extreme drought events become more frequent as predicted under global climate change[5,6]. Applying more conservative safety margins, for example, the difference between Ψ_{min} and almost complete xylem failure (Ψ_{88}), showed a similar convergence of vulnerability across forest biomes (Fig. 2b).

Overall, gymnosperms showed greater hydraulic safety margins than angiosperms, with 42% of all angiosperms versus only 6% of all gymnosperms operating at negative safety margins, that is, Ψ_{min} below Ψ_{50} (Figs 1, 2). The seemingly risky embolism-tolerance strategies seen in angiosperms suggest that flowering plants may have a greater capacity to reverse embolism, a process by which gas is dissolved and the conduits are restored to a water-filled and functional status. Although this process is still poorly understood, it is clear that recovery can only occur if periods of drought are followed by sufficient precipitation and a return to favourable water status[24,25]. Therefore, refilling does not represent an effective escape strategy for mitigating the effects of severe and persistent drought.

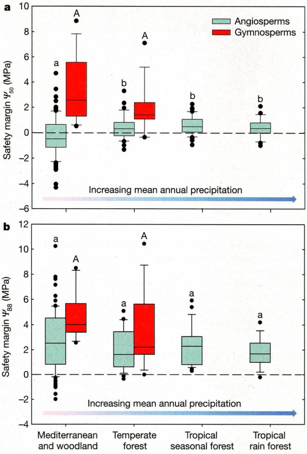


Figure 2 | Box plot of hydraulic safety margins for angiosperm and gymnosperm species across forest biomes. a, b, The Ψ_{50} ($\Psi_{min} - \Psi_{50}$) safety margin is shown in **a** ($n = 223$), and the Ψ_{88} ($\Psi_{min} - \Psi_{88}$) safety margin is shown in **b** ($n = 222$). Boxes show the median, 25th and 75th percentiles, error bars show 10th and 90th percentiles, and filled symbols show outliers. Gymnosperm species were not represented in tropical forests. Significant differences ($P < 0.05$) between biome means are indicated by letters above boxes with angiosperms (lowercase a, b) and gymnosperms (uppercase A) considered separately. Data set available in Supplementary Table 1.

Wider safety margins in gymnosperms than angiosperms do not mean that gymnosperms are immune to the threat of hydraulic failure. In fact, Pinaceae species have significantly lower embolism resistance and safety margins than Cupressaceae[19,20,26], which is reflected in the greater frequency of dieback events involving Pinaceae[1,13,26]. In the severe 2002–2003 drought, for example, *Pinus* (Pinaceae) species suffered widespread mortality in the south western United States, whereas co-existing *Juniperus* (Cupressaceae) survived[13].

In our data set, Ψ_{50} was strongly associated with MAP (Fig. 3) such that both the mean and upper tenth quantile trends showed decreasing resistance to embolism (less negative Ψ_{50}) with increasing rainfall ($P < 0.05$). Similar relationships were found between Ψ_{50} and climate variables that account for both the variation in potential evapotranspiration (PET) and seasonality of precipitation: aridity index (MAP divided by PET) and mean precipitation of the driest quarter (Supplementary Fig. 2). However, a wide range of hydraulic strategies occur within any given climate region, with the greatest variation in Ψ_{50} occurring at sites with a MAP of between 300 and 1,000 mm. In high MAP sites, represented by tropical rainforests in our data set, variation is compressed to less negative Ψ_{50}, suggesting that low embolism resistance is associated with high transport efficiency and low structural 'costs', making this an advantageous strategy in highly productive, wet tropical environments[18].

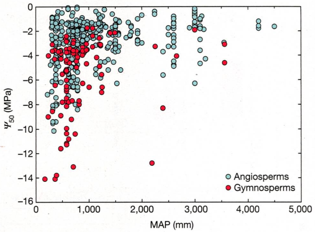


Figure 3 | Embolism resistance as a function of mean annual precipitation for 384 angiosperm and 96 gymnosperm species. Each point represents one species. A generalized model indicated that embolism resistance (Ψ_{50}) was significantly related ($P < 0.00001$) to mean annual precipitation (MAP) for angiosperms and gymnosperms (see Methods for details), with decreasing resistance to embolism corresponding to increasing rainfall. The full data set is available in Supplementary Table 1.

It is clear from Fig. 3 that Ψ_{50} and MAP are decoupled in certain cases, implying that some species growing in drier environments escape from water stress, therefore alleviating the need for high embolism resistance. There are many examples in the literature of species with low embolism resistance growing in areas of low rainfall or seasonal drought. This includes riparian and ground-water-dependent vegetation[10], and drought-deciduous trees in tropical dry forests[27]. These species avoid very negative Ψ_x by some combination of predictable access to ground water (deep roots), internal water storage and reduced leaf area or other shifts in biomass allocation[10,23,28]. Although these adjustments decouple Ψ_{50} and Ψ_{min} from MAP, it seems that the majority of species operate close to their functional limits. They are therefore exposed to xylem failure during anomalously low rainfall in a manner that is largely independent of rainfall region and biome (Fig. 2).

The convergence on a 'risky' hydraulic strategy exhibited by many species can be understood as the result of a trade-off that balances growth with protection against risk of mortality in a given environment[17,21,25]. Thus, a low safety margin to Ψ_{50} also indicates that stomatal regulation takes full advantage of the range of xylem pressures that are within the tolerance of the hydraulic system of that species. This stomatal behaviour carries with it the benefit of increased carbon gain but may come at the cost of extensive loss of photosynthetic area or death.

A fundamental question concerns the plasticity and genetic diversity of embolism resistance within species[22,29,30]. If the tight link between embolism resistance and water availability is the product of natural selection over many generations and adaptation is limited by a long generation cycle of perennial plants, then the rapid pace of climate change may outstrip the capacity of populations to adapt. This could lead to long-term reductions in net primary productivity of forest systems, loss of biodiversity and changes to the composition of forest and woodland communities. Although it is evident that multiple mechanisms (hydraulic failure, carbohydrate depletion and insect attack) are involved in drought-induced mortality, these mechanisms are highly interdependent[11,12]. Embolism formation is a key mechanism of vegetation shifts and forest decline because it sets the thresholds for stomatal closure, leading to limitations on photosynthesis, increased heat and light damage, and a run-down of carbohydrate reserves over time. Our findings demonstrate the necessity of integrating long-term monitoring of Ψ_{min} with measurements of embolism resistance and safety margins. The inclusion of these data in process-based vegetation models will improve the accuracy with which the responses of forest ecosystems to climate change can be predicted.

METHODS SUMMARY

Xylem traits were compiled from published and unpublished sources (Supplementary Table 1). When values of Ψ_{50} and Ψ_{88} were not reported in numerical form, they were extracted from published graphs of vulnerability curves. Ψ_{min} data are the minimum midday water potential recorded for each species in the field, indicating a seasonal, rather than daily minimum value. Two types of safety margins were calculated: the Ψ_{50} margin ($\Psi_{min} - \Psi_{50}$) and the Ψ_{88} margin ($\Psi_{min} - \Psi_{88}$) (Supplementary Fig. 1).

Data in Fig. 1 were binned in 1.0-MPa increments of Ψ_{50}. Bins were pooled with the next lowest bin if they contained only one sample. Regression lines in Fig. 1 presented are for raw data to avoid bias associated with uneven bin size. Differences in biome means for safety margin were analysed using a general linear model, with differences between angiosperms and gymnosperms considered separately. Climate data were taken from the paper in which Ψ_{50} data were published, or from the WorldClim database or CRU climate database, whichever gave an elevation closest to that given in the paper. Relationships between Ψ_{50} and climate variables (MAP, aridity index, mean precipitation of the driest quarter) were analysed using a generalized model assessed by restricted maximum likelihood in which variance was simultaneously modelled as a power function of the same climate variables. Quantile regression was also used to assess the boundary relationship between Ψ_{50} and climate variables. As Ψ_{50} is a negative variable, the 10% quantile was used, which is equivalent to the 90% quantile for a positive response variable.

Forest biomes were assigned based on site descriptions contained in the primary sources. We defined 'forest' broadly to include Mediterranean, savanna and woodland environments that are not commonly classified as forests. The data set therefore encompasses tree, shrub and liana species from vegetation communities with a significant component of woody plants.

Full Methods and any associated references are available in the online version of the paper.

Received 21 January; accepted 18 October 2012.

Published online 21 November 2012.

1. Allen, C. D. *et al.* A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259,** 660–684 (2010).
2. Brodribb, T. J. & Cochard, H. Hydraulic failure defines the recovery and point of death in water-stressed conifers. *Plant Physiol.* **149,** 575–584 (2009).
3. Kursar, T. A. *et al.* Tolerance to low leaf water status of tropical tree seedlings is related to drought performance and distribution. *Funct. Ecol.* **23,** 93–102 (2009).
4. McDowell, N. *et al.* Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytol.* **178,** 719–739 (2008).
5. Allison, I. *et al. The Copenhagen Diagnosis: Updating the World on the Latest Climate Science* (Elsevier, 2009).
6. Zhang, X. *et al.* Detection of human influence on twentieth-century precipitation trends. *Nature* **448,** 461–465 (2007).
7. Meir, P. & Woodward, F. I. Amazonian rain forests and drought: response and vulnerability. *New Phytol.* **187,** 553–557 (2010).
8. Phillips, O. L. *et al.* Drought sensitivity of the amazon rainforest. *Science* **323,** 1344–1347 (2009).
9. Engelbrecht, B. M. J. *et al.* Drought sensitivity shapes species distribution patterns in tropical forests. *Nature* **447,** 80–82 (2007).
10. Pockman, W. T. & Sperry, J. S. Vulnerability to xylem cavitation and the distribution of Sonoran desert vegetation. *Am. J. Bot.* **87,** 1287–1299 (2000).
11. McDowell, N. G. *et al.* The interdependence of mechanisms underlying climate-driven vegetation mortality. *Trends Ecol. Evol.* **26,** 523–532 (2011).
12. Anderegg, W. R. L. *et al.* The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. *Proc. Natl Acad. Sci. USA* **109,** 233–237 (2012).
13. Breshears, D. D. *et al.* Regional vegetation die-off in response to global-change-type drought. *Proc. Natl Acad. Sci. USA* **102,** 15144–15148 (2005).
14. Zhao, M. & Running, S. W. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* **329,** 940–943 (2010).
15. Lewis, S. L. Tropical forests and the changing earth system. *Phil. Trans. R. Soc. B* **361,** 195–210 (2006).
16. Pockman, W. T., Sperry, J. S. & O'Leary, J. W. Sustained and significant negative water pressure in xylem. *Nature* **378,** 715–716 (1995).
17. Tyree, M. T. & Sperry, J. S. Vulnerability of xylem to cavitation and embolism. *Annu. Rev. Plant Phys. Mol. Bio.* **40,** 19–38 (1989).
18. Sperry, J. S., Hacke, U. G. & Pittermann, J. Size and function in conifer tracheids and angiosperm vessels. *Am. J. Bot.* **93,** 1490–1500 (2006).
19. Maherali, H., Pockman, W. T. & Jackson, R. B. Adaptive variation in the vulnerability of woody plants to xylem cavitation. *Ecology* **85,** 2184–2199 (2004).

20. Delzon, S., Douthe, C., Sala, A. & Cochard, H. Mechanism of water-stress induced cavitation in conifers: bordered pit structure and function support the hypothesis of seal capillary-seeding. *Plant Cell Environ.* **33,** 2101–2111 (2010).
21. Sperry, J. S., Adler, F. R., Campbell, G. S. & Comstock, J. P. Limitation of plant water use by rhizosphere and xylem conductance: results from a model. *Plant Cell Environ.* **21,** 347–359 (1998).
22. Alder, N. N., Sperry, J. S. & Pockman, W. T. Root and stem xylem embolism, stomatal conductance, and leaf turgor in *Acer grandidentatum* populations along a soil moisture gradient. *Oecologia* **105,** 293–301 (1996).
23. Meinzer, F. C., Johnson, D. M., Lachenbruch, B., McCulloh, K. A. & Woodruff, D. R. Xylem hydraulic safety margins in woody plants: coordination of stomatal control of xylem tension with hydraulic capacitance. *Funct. Ecol.* **23,** 922–930 (2009).
24. Brodersen, C. R., McElrone, A. J., Choat, B., Matthews, M. A. & Shackel, K. A. The dynamics of embolism repair in xylem: *in vivo* visualizations using high-resolution computed tomography. *Plant Physiol.* **154,** 1088–1095 (2010).
25. Brodribb, T. J., Bowman, D. J. M. S., Nichols, S., Delzon, S. & Burlett, R. Xylem function and growth rate interact to determine recovery rates after exposure to extreme water deficit. *New Phytol.* **188,** 533–542 (2010).
26. Martínez-Vilalta, J., Sala, A. & Piñol, J. The hydraulic architecture of Pinaceae—a review. *Plant Ecol.* **171,** 3–13 (2004).
27. Choat, B., Ball, M. C., Luly, J. G. & Holtum, J. A. M. Hydraulic architecture of deciduous and evergreen dry rainforest tree species from north-eastern Australia. *Trees* **19,** 305–311 (2005).
28. Scholz, F. G., Phillips, N. G., Bucci, S. J., Meinzer, F. C. & Goldstein, G. in *Size- and Age-Related Changes in Tree Structure and Function* Vol. 4 (eds Meinzer, F. C. *et al*) 341–361 (Springer, 2011).
29. Lamy, J. B. *et al.* Uniform selection as a primary force reducing population genetic differentiation of cavitation resistance across a species range. *PLoS ONE* **6,** e23476 (2011).
30. Wortemann, R. *et al.* Genotypic variability and phenotypic plasticity of cavitation resistance in *Fagus sylvatica* L. across Europe. *Tree Physiol.* **31,** 1175–1182 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the ARC-NZ Vegetation Function Network for hosting the original working group from which the data set was compiled. We are grateful to the Alexander von Humboldt Foundation for supporting B.C. during preparation of the manuscript.

Author Contributions B.C. and S.J. led the initial working group and coordinated the analysis and write-up of the work. B.C., S.J., T.J.B., H.C., S.D., R.B., S.J.B., T.S.F., S.M.G., U.G.H., A.L.J., F.L., H.M., J.M.-V., S.M., M.M., P.J.M., A.N., J.P., R.B.P., J.S.S., M.W., I.J.W. and A.E.Z. contributed to compilation and organization of the data set and writing of the manuscript. S.M.G. and I.J.W. extracted climate data from the WorldClim and CRU climate databases. H.M., M.M. and J.M.-V. assisted in statistical analyses of the data set.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J. (steven.jansen@uni-ulm.de).

LETTER

doi:10.1038/nature11584

The genomic landscape of species divergence in *Ficedula* flycatchers

Hans Ellegren[1], Linnéa Smeds[1], Reto Burri[1], Pall I. Olason[1], Niclas Backström[1], Takeshi Kawakami[1], Axel Künstner[1]†, Hannu Mäkinen[1], Krystyna Nadachowska-Brzyska[1], Anna Qvarnström[2], Severin Uebbing[1] & Jochen B. W. Wolf[1]

Unravelling the genomic landscape of divergence between lineages is key to understanding speciation[1]. The naturally hybridizing collared flycatcher and pied flycatcher are important avian speciation models[2–7] that show pre- as well as postzygotic isolation[8,9]. We sequenced and assembled the 1.1-Gb flycatcher genome, physically mapped the assembly to chromosomes using a low-density linkage map[10] and re-sequenced population samples of each species. Here we show that the genomic landscape of species differentiation is highly heterogeneous with approximately 50 'divergence islands' showing up to 50-fold higher sequence divergence than the genomic background. These non-randomly distributed islands, with between one and three regions of elevated divergence per chromosome irrespective of chromosome size, are characterized by reduced levels of nucleotide diversity, skewed allele-frequency spectra, elevated levels of linkage disequilibrium and reduced proportions of shared polymorphisms in both species, indicative of parallel episodes of selection. Proximity of divergence peaks to genomic regions resistant to sequence assembly, potentially including centromeres and telomeres, indicate that complex repeat structures may drive species divergence. A much higher background level of species divergence of the Z chromosome, and a lower proportion of shared polymorphisms, indicate that sex chromosomes and autosomes are at different stages of speciation. This study provides a roadmap to the emerging field of speciation genomics.

As lineages diverge, a combination of pre- as well as postzygotic reproductive isolation barriers will eventually arise[1]. Divergence is likely to start from specific loci that may precede and cause the evolution of reproductive incompatibility. Hybridization between diverging lineages may therefore create a genomic mosaic of regions where interspecific gene flow occurs at different rates (the genic view of speciation[11]), with introgression expected to be weak in genomic regions involved in speciation. Revealing the genomic regions with elevated levels of divergence will eventually deepen our knowledge of the speciation process. However, more than 150 years after the publication of *On the Origin of Species*[12], the genetic basis of speciation is still largely unresolved[13,14]. We know little about the identity, number and effect size of loci involved in population divergence, their genomic distribution and the type of mutations involved. Advances in sequencing technology now open a promising avenue for the study of genomic divergence, even for non-model vertebrate species with gigabase (Gb)-sized genomes.

The collared flycatcher *Ficedula albicollis* and the pied flycatcher *Ficedula hypoleuca* (Fig. 1) are important study organisms for key aspects of evolutionary ecology and biology[2–7]. Diverged less than 2 million years ago, their history has been shaped by repeated cycles of glaciation in Eurasia where periods of allopatric divergence in refugia probably alternated with periods of secondary contact during which gene flow and selection were vital components; they still hybridize in areas of sympatry (Supplementary Figure 2). To study the genetic basis of species divergence in this system, we sequenced and assembled the flycatcher genome, and physically placed, ordered and oriented sequence scaffolds along chromosomes through linkage-map data. This was followed by re-sequencing of genomes and transcriptomes of population samples of both species (Supplementary Methods, Supplementary Fig. 1), allowing base-pair (bp)-resolution of the pattern of differentiation on a genomic level and providing a roadmap for studies in the emerging field of speciation genomics.

The final assembly encompassed 1.13 Gb with an N50 scaffold size of 7.3 Mb and with 89% of the assembly contained within no more than 200 scaffolds larger than 1 Mb (Supplementary Tables 1–7). The sequenced bird was heterozygous at 3.66 million positions, corresponding to an average of one segregating site every 330 bp. A low-density linkage map of collared flycatcher[10] anchors 73% of the assembly to chromosomes and orients scaffolds along them (Supplementary Fig. 3). Based on conserved chromosomal organization between flycatcher and zebra finch (Supplementary Fig. 4) we were able to anchor and orient additional scaffolds, thereby physically positioning 1.00 Gb of the assembly (89%) in the genome (Supplementary Fig. 5). This illustrates that physical assembly of Gb-sized genomes sequenced with short reads is possible with modest linkage information and when assisted with genome information from a related species. The flycatcher genome contained 18,735 predicted protein-coding genes, of which 18,649 (>99.5%) were identified as expressed based on RNA–seq data from a variety of tissues.

We then sequenced the genomes of 10 unrelated males of each species (mean coverage 5.69x ± 2.01 s.d.; Supplementary Table 8) and found 9.86- and 7.13-million segregating sites in collared flycatchers and pied flycatchers, respectively. The fact that 3.81 million of these single nucleotide polymorphisms (SNPs) (53.4% and 38.6% of the total in each species, respectively) were shared between species confirms their close genetic relationship and provides an unusual access to genomic data of two species before complete lineage sorting. The mean pairwise nucleotide difference in interspecific comparisons (d_{xy}) for

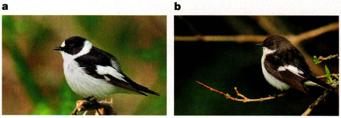


Figure 1 | **Study species.** **a**, Male collared flycatcher. **b**, Male pied flycatcher. Note that the male collared flycatcher has a white neck collar and a more pronounced white forehead and wing patches. Photographs courtesy of Johan Träff.

---

[1]Dept of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. [2]Dept of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. †Present address: Max Planck Institute for Developmental Biology Department of Molecular Biology (VI), Spemannstrasse 37-39, D-72076 Tübingen, Germany.

50-kb windows was 0.0046 ± 0.0011, which was only marginally higher than the mean pairwise nucleotide diversity (π) in intraspecific comparisons (π_{coll}: 0.0036 ± 0.0010; π_{pied}: 0.0021 ± 0.00076); individuals of the two species are thus genetically not much more different from each other than are individuals within species. Acknowledging that population samples of 10 individuals from each species provide low power for detecting rare alleles, an indication of species divergence can be obtained by noting at how many sites all collared flycatchers were homozygous for one allele and all pied flycatchers homozygous for another (which we refer to as sites of fixed differences, d_f). We found 239,745 such sites, which corresponds to 1 fixed difference every 4.7 kb. Of these, 1,513 sites were located within protein-coding regions and it is realistic that a proportion of these positions represent the genetic basis for key phenotypic differences between the two species.

The genomic landscape of species divergence was highly heterogeneous, with a fraction of windows showing highly elevated divergence up to 50 times higher than the genomic median (0.00013) and mean (0.00022) (Supplementary Fig. 6). The distribution of autosomal windows with elevated divergence showed a non-random pattern with approximately 50 well-defined clusters of high F_{ST} (the fixation index, a measure of population differentiation) and d_f ('divergence peaks' or 'genomic islands of divergence') (Fig. 2a and Table 1). Average peak size was in the range of several hundred kb (median, 400 kb; mean, 625 kb; range, ~100 kb to 3 Mb; Supplementary Fig. 7) and, in total, divergence peaks covered 2.7% of the genome yet containing 25% of all fixed differences. An immediate feature of the distribution of these 50 peaks was that they were non-randomly distributed across the genome (Kolmogorov–Smirnov test, $D = 0.4516$, $P = 0.0002$), irrespective of chromosome size and despite substantial heterogeneity in chromosome size, there were in most cases one to three peaks per chromosome (Fig. 2a). Moreover, peaks were highly overrepresented in the very end of chromosomes and six microchromosomes had peaks in both ends. Another feature of divergence islands was that they lie at the end of scaffolds, thereby not forming a continuous and symmetric signal in the assembly. As a consequence, peaks in the end of chromosomes were generally 'one-tailed', whereas peaks within the interior of chromosomes were usually formed by adjacent scaffolds with maximum divergence juxtaposed to the assembly gap between scaffolds (Fig. 2b and Supplementary Fig. 8). This peculiar pattern raises the issue of whether peaks are artefacts associated with the scaffolding process or read mapping. However, several observations convincingly argue against this (Supplementary Notes).

Table 1 | Mean values of population genomic parameters

Parameter	Genomic background	Islands of divergence	Extreme per peak
d_f	0.00027	0.00171	0.00281
F_{ST}	0.357	0.742	0.856
π_{pied}	0.00219	0.00067	0.00065
π_{coll}	0.00370	0.00132	0.00030
D_{pied}	0.376	0.129	−0.428
D_{coll}	0.221	0.053	−0.466
r^2_{pied}	0.059	0.082	0.133
r^2_{coll}	0.065	0.088	0.111

Data are from autosomal 50-kb windows divided into the genomic background and genomic islands of divergence (windows with the density of fixed differences, $d_f > 0.001$). D, Tajima's D, r^2, an estimate of linkage disequilibrium. 'Extreme per peak' represents the mean of the highest or lowest value per peak. Differences between all parameter estimates in divergence islands versus genomic background are statistically significant at $P < 2.2 \times 10^{-16}$ (Wilcoxon test).

If selection has driven population divergence in regions of high differentiation, we might expect to see reduced levels of within-species diversity in these regions in one of the species. Species-specific estimates of π showed that this was essentially always the case (Fig. 2b, Supplementary Fig. 8), with mean π in divergence islands less than one-third of the genomic background level (Table 1). As F_{ST} by its nature is a function of within-species diversity[15], this association might be trivial. However, because the vast majority of divergence islands were seen with F_{ST} as well as d_f, regions of high divergence were characterized both by a high frequency of sequence differences between the two species and a low frequency of sequence differences within species. A noteworthy consequence of these coinciding features was that d_{xy} did not exceed background levels in divergence islands (Fig. 2b and Supplementary Fig. 8). Further indication of selection in

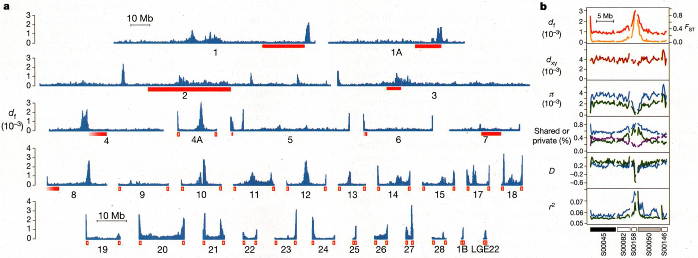


Figure 2 | The genomic landscape of species divergence in flycatchers. **a**, Distribution of divergence measured as the density of fixed differences per bp for 200-kb windows across the genome. Chromosomes are listed in numerical order and are separated by gaps. Red horizontal bars show the approximate location of centromeres in homologous chromosomes of zebra finch. Open read symbols are used to indicate that avian microchromosomes are generally acro- or telocentric; both ends of these chromosomes are labelled as the orientation is not known. For chromosomes 4, 6 and 8, there is a lack of an *in situ* mapped marker 5′ of the centromere in zebra finch. **b**, Distribution of population genomic parameters along an example chromosome (chromosome 4A). The plots show the density of fixed differences per bp (d_f) (yellow), F_{ST} (red), the total between-species sequence divergence (d_{xy}), nucleotide diversity (π) for each species, the proportion of shared polymorphisms among sites polymorphic in at least one species (purple), the proportion of private polymorphisms among sites polymorphic within species (private and shared polymorphisms shown in the same panel), Tajima's D, and linkage disequilibrium (r^2). For π, private polymorphisms, D and r^2, species-specific estimates are given for collared flycatcher in blue and for pied flycatcher in green. Assigned scaffolds are shown under the plot: black, denoting scaffolds ordered and oriented by the collared flycatcher linkage map; grey, scaffolds ordered with the collared flycatcher linkage map and oriented through comparative mapping with zebra finch; white, scaffolds ordered and oriented through comparative mapping with zebra finch.

regions of high divergence was given by the observations of allele-frequency spectra being skewed towards rare alleles and strong signals of linkage disequilibrium (Table 1, Fig. 2b and Supplementary Notes).

How the abovementioned signs of selection are distributed between the two species is relevant for interpretation of the heterogeneous genomic landscape of species divergence. For almost all regions of elevated divergence, both species showed reduced nucleotide diversity (Fig. 2b and Supplementary Fig. 8). This cannot be explained by a loss of diversity in the ancestral population because it would not lead to the observed high incidence of fixed differences. Moreover, taking d_S (the synonymous substitution rate) as a proxy for mutation rate, we found that regions of high divergence were not low in variability because of low mutation rate (generalized linear model, including chromosome length, peak versus non-peak; $z = 0.598$, $P = 0.550$). The association of high divergence with low diversity speaks further against mapping artefacts, which can increase divergence between species but should also lead to elevated diversity within species. Taken together, these results suggest that selection has acted to reduce genetic variability in the very same regions in the two lineages independently.

Genomic regions with reduced levels of interspecific recombination are hindered from gene flow, facilitating the build-up of reproductive incompatibilities. Moreover, divergent selection may enhance differentiation over larger genomic regions when the intraspecific recombination rate is low (divergence hitchhiking[16,17]). To assess the relationship between recombination and divergence, we estimated the population recombination rate (ρ, which equals $N_e r$, where N_e is the effective population size and r is the per-generation recombination rate), ρ/π (as reduced diversity, N_e, within divergence islands will lower the estimates of ρ and contribute to differences in ρ between divergence peaks and the genomic background even if r would be similar) and used genetic distances from the collared flycatcher linkage map related to the physical distance between markers according to the genome assembly to assess the relationship between recombination and divergence (Supplementary Methods and Supplementary Table 9). These tests provided no strong evidence of reduced recombination rate in the proximity of divergence islands (Supplementary Notes).

The flycatcher karyotype has not been established, thus the location of centromeres is not known. However, avian microchromosomes are generally acro- or telocentric[18]. An attempt to approximate the location of centromeres on flycatcher macrochromosomes was made using information on the location of centromeres in the karyotype of zebra finch, coupled with the high degree of flycatcher–zebra-finch synteny conservation[10] (Supplementary Fig. 4). This reveals considerable overlap between the presumed location of flycatcher centromeres and divergence islands (Fig. 2a). This is particularly apparent when considering the enrichment of divergence islands at the ends of microchromosomes. Moreover, the fact that several microchromosomes showed divergence peaks in both ends further suggests that at least some telomeric regions are highly differentiated between species.

Limited pedigree data from multiple generations of flycatcher hybrid descendants demonstrates fitness reduction and suggests that the current rate of introgression is low[9]. Nevertheless, genetic data from a few loci have previously indicated detectable levels of ongoing gene flow[8]. To further address this issue we carried out deep sequencing of 24 intronic regions spread across the genome in sympatric population samples of the 2 species (Supplementary Methods and Supplementary Table 10). Assuming an isolation–migration model, the maximum likelihood estimate of the rate of gene flow from pied flycatcher to collared flycatcher was 1.7×10^{-6} per gene and generation (90% posterior density distribution 0.1–2.8×10^{-6}), while the rate for the opposite direction was much lower (4.5×10^{-9}). From the analysis of nested models we can reject a model without gene flow from pied flycatcher to collared flycatcher (likelihood ratio test, $P < 0.01$) and estimate the rate at $N_e m = 0.38$ (where m is the migration rate), or roughly one migrant every three generations.

If divergence islands are involved in reproductive isolation, they might be expected to be shielded from gene flow in areas of sympatry, while other genome regions may get introgressed. One way of addressing this possibility is to study the distribution of private and shared polymorphisms to infer differential rates of lineage sorting, such as those caused by variation in gene flow across the genome. We found a very clear pattern at the point at which the proportion of shared polymorphisms drops significantly in all divergence islands, from a mean background level of 32.8% to 18.3% in islands (Wilcoxon test, $W = 632{,}244$, $P \ll 0.001$; Fig. 2b and Supplementary Fig. 8). This observation indicates more advanced lineage sorting within than outside regions of elevated divergence, and is consistent with a role for gene flow in homogenizing background levels of divergence. Moreover, we found that the proportion of private polymorphisms was significantly higher in divergence islands than elsewhere in the genome (pied flycatcher: 35.3% in non-islands versus 56.1% in islands, $W = 319{,}712$, $P \ll 0.001$; collared flycatcher: 60.2% in non- islands vs. 75.4% in islands, $W = 608{,}994$, $P \ll 0.001$; Fig. 2b and Supplementary Fig. 8), consistent with restricted gene flow in islands. Furthermore, the genomic background level of the proportion of private polymorphisms was considerably higher in collared flycatcher than in pied flycatcher ($W = 19{,}001{,}201$, $P \ll 0.001$, Fig. 2b and Supplementary Fig. 8), in agreement with the direction of gene flow recorded.

Birds have female heterogamety (males, ZZ; females, ZW) and as all sequencing was carried out using male birds, read coverage is expected to be similar for autosomes and the Z chromosome. Estimates of diversity and divergence should therefore be directly comparable between chromosome categories. The Z chromosome showed greater than sevenfold higher mean divergence (d_f, 0.0016 ± 0.00060) than autosomes (0.00022 ± 0.00036, $W = 23{,}706{,}977$, $P < 2 \times 10^{-16}$) and significantly higher F_{ST} (0.623 ± 0.076 versus 0.350 ± 0.110, $W = 23{,}274{,}298$, $P < 2 \times 10^{-16}$) (Fig. 3). Divergence was more uniformly distributed along the Z chromosome and did not show the distinct islands of divergence characterizing most autosomes; the Z chromosome contained approximately 35% of all fixed sites in the genome. Moreover, estimates of π, Tajima's D and r^2 (an estimate of linkage disequilibrium) were also more uniform along the Z chromosome (Fig. 3 and Supplementary Fig. 8). Despite the higher mean divergence, we note that divergence at individual windows on the Z chromosome did not exceed that within autosomal divergence islands. High sex-linked divergence is thus a consequence of increased background level rather than more extreme divergence in individual regions. Reports of a disproportionately large effect of the X chromosome on hybrid sterility and of reduced introgression of sex-linked genes[19] have fed the idea that sex chromosomes are particularly

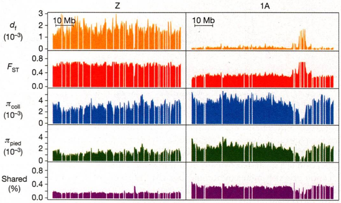


Figure 3 | Contrasting levels of divergence and diversity between the Z chromosome and a similarly sized autosome (chromosome 1A). The plots show the density of fixed differences per kb (d_f), F_{ST}, nucleotide diversity (π; collared flycatcher in blue and pied flycatcher in green) and the proportion of shared polymorphisms.

important for the build-up of reproductive isolation. This is supported by data from female heterogametic organisms[20], including higher F_{ST} seen for a handful of Z-linked markers than for autosomal markers in flycatchers[21]. In flycatchers, mating patterns also suggest sex-linkage of male plumage traits and species recognition[5,21], traits that may evolve under the influence of divergent selection. Our observations could be taken to suggest that natural or sexual selection at multiple loci associated with reproductive isolation on the Z chromosome has erased the signal from individual divergence islands by broadly increasing divergence to a higher background level. From this perspective, the Z chromosome could be seen to represent a more advanced stage of species differentiation, with islands turning into plateaus or divergence hitchhiking turning into genome hitchhiking[22]. We note in this context that total sequence differentiation in interspecific comparisons was higher for the Z chromosome (mean $d_{xy} = 0.0057 \pm 0.0011$) than for autosomes ($0.0045 \pm 0.0010$; $W = 19{,}256{,}489$, $P < 2 \times 10^{-16}$). Moreover, more advanced divergence of the Z chromosome compared to autosomes was also supported by a significantly lower proportion of shared polymorphisms in the former (15.2%) than among the latter (32.3%; $W = 1{,}408{,}692$, $P \ll 0.001$) (Fig. 3).

The 50 regions defined as divergence peaks contained a total of 530 protein-coding genes. An assessment of gene ontology among these genes did not reveal any functional category to be significantly over-represented (Supplementary Table 11). Moreover, we found no indication that proteins encoded by genes within divergence peaks would be faster evolving than other proteins in the genome (general linearized model of d_N/d_S, where d_N is the rate of non-synonymous substitution, including chromosome length; peak versus non-peak; $z = 0.837$, $P = 0.403$). As many peak regions contained more than one gene it is possible that unrelated features of linked genes blur characteristics common to genes under selection. However, we found that genes differentially expressed between species were significantly more common in peak regions (246 out of 346 genes; 71.1%) than in the rest of the genome (4,180 out of 7,134 genes, 58.6%; $\chi^2 = 11.2$, $P < 10^{-3}$; Supplementary Table 12). One possible explanation to this observation is that standing variation at *cis*-acting regulatory elements in an ancestral population has segregated via linkage to loci under divergent selection in peak regions.

The collared flycatcher and the pied flycatcher probably started to diverge in allopatry in glacial refugia of the Mediterranean area during the Pleistocene epoch, candidate regions being the Iberian and Apennine peninsulas, respectively. Subsequent secondary contact during repeated cycles of interglacial periods allowed gene flow, just as hybridization and gene flow occurs in contemporary areas of sympatry. According to this scenario, allopatric divergence may have been followed repeatedly by genomic homogenization in sympatry. The highly heterogeneous nature of genomic divergence between the two species is compatible with such a model, with some genomic regions refractory to gene flow. Our data show that these regions are localized, numerous, diverged far beyond the background level and present on almost all chromosomes, essentially shedding light on several of the central questions on the genomic landscape of species divergence. The consistent observation in both flycatcher species of reduced diversity in divergence islands would suggest that the same loci, or closely linked loci, have been subject to directional selection in both lineages independently. This, together with the juxtaposition of the most extreme divergence and gaps in the genome assembly, raise the possibility that centromeres or other heterochromatic repeats themselves actually are drivers of species divergence. The meiotic drive model of speciation invokes an arms race between centromeric alleles for deposition into the single resultant oocyte of female meiosis[23], in which selection acts on allelic variation in the ability to attract microtubuli of an asymmetric spindle pole. This could lead to rapid evolution of repeat sequences as well as of proteins involved with spindle-fibre attachment to centromeres, possibilities that are supported by empirical data[24], and may hinder proper chromosome segregation or pairing during hybrid meiosis[25]. Interestingly, the *Drosophila Zhr* locus causing female lethality is itself a heterochromatic satellite-DNA block[26], and the *Drosophila* ods-site homeobox (*OdsH*) speciation gene has been linked to hybrid male sterility through its binding to evolutionary labile heterochromatic repeats[27]. Similar actions are suggested for other speciation genes[13]. There is evidence for segregation distortion in chicken chromosomes involving loci with centromeric or telomeric locations[28]. The genomic distribution of divergence peaks in flycatchers is compatible with an involvement of telomeres as well, as some chromosomes showed divergence signals in both ends. Telomeres have been shown to have an evolutionary conserved role during meiosis in which they cluster on the nuclear envelope, forming a 'telomere bouquet', and enable chromosome movements to promote homologous synapsis[29]. It is noteworthy in this respect that birds have more extensive arrays of telomeric repeats than other vertebrates, and show structural polymorphism of telomeres within species[30], setting the stage for a meiotic drive also in these types of repeats. As meiotic drives are characterized by repeated episodes of selection, this would be compatible with the relative large size of divergence islands observed.

To conclude, this study presents the genome sequence of an avian speciation model and unravels the genomic landscape of species divergence in unprecedented detail. The results show strong heterogeneity in sequence differentiation in a species pair in which lineage sorting is incomplete. The potential connection of species divergence to key repetitive elements of chromosomes calls for a shift in focus, with the quest for genetic basis of reproductive isolation extended to include sequences other than protein-coding genes. For further dissection of the mechanism driving species divergence in this and other systems it will be important to obtain detailed maps of how the rate of recombination varies along chromosomes, based on large-scale genotyping in pedigrees. Together with modelling (under varying intensity and character of selection), this can address to what extent sweeps or background selection, possibly aided by low recombination, are expected to increase divergence and over what distances. As the size of genomic islands of divergence will also be affected by variation in N_e and the rate of migration, these are also factors that need to be integrated in models. Moreover, extensive genotyping in pedigrees would be a means to test for segregation distortion introduced by meiotic drives.

METHODS SUMMARY

Genome sequencing was carried out with Illumina technology using DNA from a single wild-caught male collared flycatcher. Sequences from paired-end and mate-pair reads of multiple libraries (200–21,000 bp) were assembled using SOAPDENOVO in subsequent steps with increasing insert size of libraries. Scaffolds were physically anchored to chromosomes by the aid of a collared-flycatcher linkage map and with comparative map information from the zebra-finch genome. Protein-coding genes of the flycatcher genome were retrieved through a combination of mapping reads to zebra-finch gene templates, using flycatcher expressed sequence tag (EST) evidence and *ab initio* prediction. Levels of gene expression were measured across a suite of tissues (embryonic, adult somatic and gonadal tissues) using RNA–seq with Illumina technology, and differentially expressed genes were identified with BAYSEQ. Population genomic analyses were based on data from re-sequencing of 10 individuals each of collared flycatcher and pied flycatcher, in which reads were mapped to the assembly using BWA (Burrows-Wheeler Aligner) software. After analysis and data processing using a combination of software tools, sequence variants were identified with the Genome Analysis Toolkit (GATK; Broad Institute). Divergence and diversity parameters were estimated using 'haploidized' data by randomly choosing one allele from heterozygous genotypes.

Received 3 April; accepted 12 September 2012.

Published online 24 October 2012.

1. Coyne, J. A. & Orr, H. A. *Speciation.* (Sinauer Associates, 2004).
2. Ellegren, H., Gustafsson, L. & Sheldon, B. C. Sex ratio adjustment in relation to paternal attractiveness in a wild bird population. *Proc. Natl Acad. Sci. USA* **93,** 11723–11728 (1996).
3. Saetre, G.-P. *et al.* A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* **387,** 589–592 (1997).

4. Qvarnström, A., Part, T. & Sheldon, B. C. Adaptive plasticity in mate preference linked to differences in reproductive effort. *Nature* **405,** 344–347 (2000).
5. Veen, T. *et al.* Hybridization and adaptive mate choice in flycatchers. *Nature* **411,** 45–50 (2001).
6. Merilä, J., Kruuk, L. E. B. & Sheldon, B. C. Cryptic evolution in a wild bird population. *Nature* **412,** 76–79 (2001).
7. Saether, S. A. *et al.* Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* **318,** 95–97 (2007).
8. Borge, T., Lindroos, K., Nadvornik, P., Syvanen, A. C. & Saetre, G. P. Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. *J. Evol. Biol.* **18,** 1416–1424 (2005).
9. Wiley, C., Qvarnstrom, A., Andersson, G., Borge, T. & Saetre, G. P. Postzygotic isolation over multiple generations of hybrid descendens in a natural hybrid zone: how well do single-generation estimates reflect reproductive isolation? *Evolution* **63,** 1731–1739 (2009).
10. Backström, N. *et al.* A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics* **179,** 1479–1495 (2008).
11. Wu, C.-I. The genic view of the process of speciation. *J. Evol. Biol.* **14,** 851–865 (2001).
12. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* (John Murray, 1859).
13. Presgraves, D. C. The molecular evolutionary basis of species formation. *Nature Rev. Genet.* **11,** 175–180 (2010).
14. Nosil, P. & Schluter, D. The genes underlying the process of speciation. *Trends Ecol. Evol.* **26,** 160–167 (2011).
15. Charlesworth, B. Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* **15,** 538–543 (1998).
16. Feder, J. L. & Nosil, P. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64,** 1729–1747 (2010).
17. Via, S. & West, J. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17,** 4334–4345 (2008).
18. Shields, G. F. Comparative avian cytogenetics: a review. *Condor* **84,** 45–58 (1982).
19. Coyne, J. A. Genetics and speciation. *Nature* **355,** 511–515 (1992).
20. Jiggins, C. D. *et al.* Sex-linked hybrid sterility in a butterfly. *Evolution* **55,** 1631–1638 (2001).
21. Saetre, G.-P. *et al.* Sex chromosome evolution and speciation in Ficedula flycatchers. *Proc. R. Soc. Lond. B* **270,** 53–59 (2003).
22. Nosil, P. & Feder, J. L. Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B* **367,** 332–342 (2012).
23. Henikoff, S. & Malik, H. S. Centromeres: selfish drivers. *Nature* **417,** 227 (2002).
24. Malik, H. S. & Henikoff, S. Major evolutionary transitions in centromere complexity. *Cell* **138,** 1067–1082 (2009).
25. Fishman, L. & Saunders, A. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* **322,** 1559–1562 (2008).
26. Ferree, P. M. & Barbash, D. A. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* **7,** e1000234 (2009).
27. Bayes, J. J. & Malik, H. S. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* **326,** 1538–1541 (2009).
28. Axelsson, E. *et al.* Segregation distortion in chicken and the evolutionary consequences of female meiotic drive in birds. *Heredity* **105,** 290–298 (2010).
29. Tsai, J.-H. & McKee, B. D. Homologous pairing and the role of pairing centers in meiosis. *J. Cell Sci.* **124,** 1955–1963 (2011).
30. Delany, M. E., Gessaro, T. M., Rodrigue, K. L. & Daniels, L. M. Chromosomal mapping of chicken mega-telomere arrays to GGA9, 16, 28 and W using a cytogenomic approach. *Cytogenet. Genome Res.* **117,** 54–63 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements Financial support was obtained from a European Research Council Advanced Investigator Grant (NEXTGENMOLECOL), a Knut and Alice Wallenberg Scholar Grant, and from the Swedish Research Council to H.E. R.B. was funded by the Swiss National Science Foundation (grants PBLAB3-134299 and PBLAB1-140171). We are grateful to M. Lascoux, M. Noor and T. Slotte for helpful discussion and comments. We thank the Uppsala University SNP and SEQ Technology Platform for help with DNA sequencing, and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), and the associated Next generation sequencing Cluster and Storage (UPPNEX) project, funded by the Knut and Alice Wallenberg Foundation and the Swedish National Infrastructure for Computing (SNIC), for computer resources.

Author Contributions L.S., P.I.O. and A.K. carried out the bioinformatic analyses; R.B. and T.K. performed population genomic analyses and interpreted the data; H.E., N.B., K.N.-B. and J.B.W.W. interpreted the data; S.U. performed analyses of differential gene expression; H.M. generated the RNA-seq data; N.B. collected and processed the samples; A.Q. facilitated sampling of collared flycatchers; H.E., L.S., R.B., P.I.O., T.K., A.K. and S.U. wrote the Supplementary Information, with input from the other authors; H.E. conceived and designed the study, supervised the project, and wrote the main paper with input from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike license, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.E. (Hans.Ellegren@ebc.uu.se). The flycatcher genome sequence and associated data is available under the accession numbers AGTO00100000 (GenBank) and ERP001377 (Short Read Archive).

doi:10.1038/nature11587

A map of visual space in the primate entorhinal cortex

Nathaniel J. Killian[1,2], Michael J. Jutras[1] & Elizabeth A. Buffalo[1,3]

Place-modulated activity among neurons in the hippocampal formation presents a means to organize contextual information in the service of memory formation and recall[1,2]. One particular spatial representation, that of grid cells, has been observed in the entorhinal cortex (EC) of rats and bats[3–5], but has yet to be described in single units in primates. Here we examined spatial representations in the EC of head-fixed monkeys performing a free-viewing visual memory task[6,7]. Individual neurons were identified in the primate EC that emitted action potentials when the monkey fixated multiple discrete locations in the visual field in each of many sequentially presented complex images. These firing fields possessed spatial periodicity similar to a triangular tiling with a corresponding well-defined hexagonal structure in the spatial autocorrelation. Further, these neurons showed theta-band oscillatory activity and changing spatial scale as a function of distance from the rhinal sulcus, which is consistent with previous findings in rodents[4,8–10]. These spatial representations may provide a framework to anchor the encoding of stimulus content in a complex visual scene. Together, our results provide a direct demonstration of grid cells in the primate and suggest that EC neurons encode space during visual exploration, even without locomotion.

The primate hippocampal formation has been shown to represent both egocentric spatial information, with cells responding to self-motion and head direction, as well as allocentric spatial information, with cells firing more spikes when a particular region of an environment is viewed[1,11–14]. The presence of allocentric spatial view cells in the hippocampus, analogous to rodent place cells, suggests that the primate entorhinal cortex (EC) may also represent space independent of the position of the animal. Indeed, the non-retinocentric, bilateral receptive fields of primate EC neurons allow for the possibility of regular and allocentric spatial firing within the EC[15]. Furthermore, a recent study has provided evidence for grid cells based on human neuronal population responses in the EC that are periodic with respect to position in virtual space[16]. To examine spatial representations in the primate hippocampal formation, we recorded the activity of 342 neurons from the EC and hippocampus of three monkeys performing a free-viewing visual recognition memory task, the visual preferential looking task (VPLT)[6,7]. Novel images were each presented twice on a computer monitor with a fixed reference frame, and gaze location and neuronal data were recorded simultaneously. Images consisted of photographs that contained a wide variety of elements, including abstract art, animals, landscapes and people, and the monkeys explored the static images with a dynamic sequence of fixations (Fig. 1a). On average, EC neurons gave enhanced responses to stimulus presentation and many demonstrated a reduced firing rate for repeat presentations, consistent with previous findings of match suppression (Fig. 2)[15]. Neurons in the hippocampus exhibited a more diverse range of suppressed and excited visual and recognition responses[6]. All of these responses were contained within the analysed data; we included data segments in the analyses only if the monkey was actively exploring an image on the screen (pre-stimulus fixation periods and viewing outside of the image bounds were not included).

In the rat EC, grid cells are found exclusively in the medial EC (MEC). Spatial and visual input to the EC is topographically organized such that spatial information reaches the MEC through input from the postrhinal cortex (the parahippocampal cortex in primates), while object information reaches lateral EC (LEC) via the perirhinal cortex[17]. In monkeys, roughly the posterior half of the EC receives visuospatial information from parahippocampal cortex, retrosplenial cortex and presubiculum, while the anterior half receives visual object information from the perirhinal cortex[18]. Projections from EC to the hippocampus also have a similar anatomical organization in rats and monkeys, but the anatomical axes are oriented differently. The anterior EC in the monkey and LEC in the rat project to the region around the border between CA1 and subiculum, whereas the posterior EC in the monkey and MEC in the rat project to proximal CA1 and distal subiculum[19]. Based on the topography of these inputs, we would expect the rat MEC and LEC to correspond to the monkey posterior and anterior EC, respectively.

The spatial firing fields of grid cells can be thought of as representing the nodes of a triangular grid, and it follows that the spatial autocorrelation of the firing-rate map should have a well-defined hexagon surrounding a central peak. A gridness score has been used to quantify the strength of 60-degree rotational symmetry that leads to this hexagonal structure (Supplementary Fig. 1)[10,20]. The gridness of all recorded units was evaluated and surrogate data were used to determine significant gridness scores. Grid cells were classified as those having a gridness score above the 95th percentile of the scores for 100 time-shifted permutations of the spike timings (Fig. 1b, e). The number of grid cells identified in the posterior EC (23 out of 193; 11.9%) was significantly greater than that expected by chance ($P < 0.0001$, $\chi^2(19.4,1)$) (Fig. 1b, c; see also Supplementary Figs 2a and 3). Importantly, the spatial density of gaze location did not produce high gridness scores (Supplementary Fig. 2b), and across the population, the grid scores for EC grid cells were consistently higher than grid scores calculated for eye movements ($P < 1 \times 10^{-8}$, Wilcoxon rank-sum test; Supplementary Fig. 2b). Because these analyses were computed across presentations of multiple complex stimuli, these data suggest that the grid-cell representation is not specific to stimulus content. To address this explicitly, we examined the reliability of these representations by comparing the rate maps generated from the first and second halves of each session. There was a significant positive correlation between these firing rate maps across the population of grid cells ($P < 1 \times 10^{-5}$, $n = 23$, Wilcoxon signed-rank test), which suggests that these representations are stable and stimulus-independent (see Supplementary Fig. 4).

The proportion of grid cells identified in the hippocampus (4 out of 119; 3.4%) was not significantly different from chance ($P = 0.41$, $\chi^2(0.67,1)$) and was significantly less than the number of EC grid cells ($P = 0.009$, $\chi^2(6.82,1)$) (Fig. 1c). An important caveat is that our hippocampal recordings were all taken from the anterior hippocampus[7], which is presumably comparable to the temporal hippocampus in rats, containing cells with larger place fields than in septal regions. Accordingly, it is difficult to compare the scale of the spatial representations between the EC and the hippocampus with the present data.

We also found a significant population of neurons in the EC, but not the hippocampus, with increased firing rate near the edges of the stimuli, quantified using a border score as described in ref. 5 (Fig. 1c, e) (EC, $P = 0.0058$, $\chi^2(7.61,1)$, 18 out of 193 (9.3%); hippocampus,

---

[1]Yerkes National Primate Research Center, 954 Gatewood Road, Atlanta, Georgia 30329, USA. [2]Wallace H. Coulter Department of Biomedical Engineering at the Georgia Institute of Technology and Emory University, 313 Ferst Drive, Atlanta, Georgia 30332, USA. [3]Department of Neurology, Emory University School of Medicine, 1440 Clifton Road, Atlanta, Georgia 30322, USA.

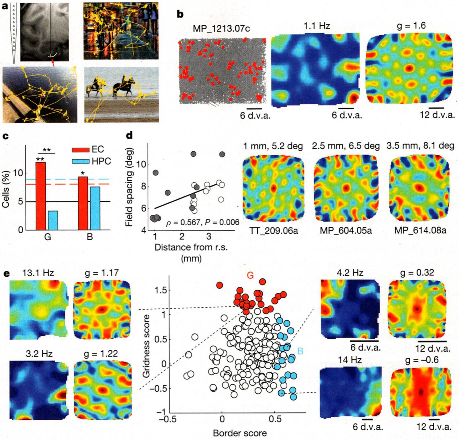


Figure 1 | Spatial representation in the primate entorhinal cortex. **a**, Recordings were carried out using a linear electrode array placed in the entorhinal cortex (red arrow). Three example 10-s scan paths are shown in yellow. **b**, An example of an entorhinal grid cell. Left, plots of eye position (grey) and spikes (red) reveal non-uniform spatial density of spiking. For clarity, only spikes corresponding to locations of firing rate above half of the mean rate were plotted. The monkey's name and unit number are indicated at the top. Middle, spatial firing-rate maps show multiple distinct firing fields. Maps are colour coded from low (blue) to high (red) firing rates. The maximum firing rate of the map is indicated at the top. Right, the spatial periodicity of the firing fields shown against spatial autocorrelations. The colour scale limits are ± 1 (blue to red), with green being 0 correlation. d.v.a., degrees of visual angle; g, gridness score. **c**, Percentages of cells in the EC and hippocampus (HPC) with a significant gridness score (G) or border score (B). The black line shows the 5% chance level, the dashed lines represent the 95% confidence level. $*P < 0.05$, $**P < 0.01$. **d**, Left, grid-cell spacing increased with distance from the rhinal sulcus (r.s.), consistent with a dorsal–ventral gradient in rodents and bats. Open and closed circles identify the grid cells from each of the two monkeys. Right, autocorrelations for representative grid cells recorded at different locations medial to the rhinal sulcus. The monkeys' names and unit numbers are indicated at the bottom. **e**, Gridness and border scores are plotted for all cells recorded in the posterior EC ($n = 193$; red, cells with significant gridness scores, $n = 23$; blue, cells with significant border scores, $n = 18$).

$P = 0.1995$, $\chi^2(1.65,1)$, 9 out of 119 (7.6%); see Supplementary Information for a detailed description of methods). The presence of cells that represent spatial locations relative to the stimulus bounds independent of stimulus content would suggest that representations of objects within an environment may be anchored to a consistent framework, with the stimulus bounds serving as landmarks. However, it is possible that these neurons simply have oblong view fields near one or more edges of the image-presentation region. This would need to be examined in future studies with changing image frames.

In the rat and bat EC, grid cells increase in field spacing in the dorsomedial to ventrolateral direction, as the distance from the border of the MEC with the postrhinal cortex increases[3,4]. This gradient mirrors the increase in the size of hippocampal place fields along the dorsoventral axis[21], to which the MEC provides input in a topographical manner[17,22]. In the monkey, cells located in lateral EC (close to the rhinal sulcus) project to posterior levels of the hippocampus, whereas cells located in the medial EC project to more anterior levels of the hippocampus[19,22]. In the present study, firing-field spacing was significantly correlated with the distance from the rhinal sulcus for each monkey, and for both monkeys together (Spearman's rank correlation coefficient: $\rho = 0.671$, $P = 0.024$ for monkey MP; $\rho = 0.665$, $P = 0.026$ for monkey TT; $\rho = 0.567$, $P = 0.006$ for both monkeys) (Fig. 1d; see also Supplementary Fig. 5).

In addition to spatial representations, we examined neuronal responses in the EC that might underlie recognition memory. Consistent with the input from perirhinal cortex to anterior EC, we found that neurons in anterior EC showed stronger visual and memory responses than did neurons in posterior EC. Specifically, neurons recorded at more anterior locations were more likely to be responsive to visual stimuli (likelihood ratio test, $P = 0.0062$, $\chi^2(7.5,1)$), and at progressively more anterior locations, larger proportions of these visually responsive neurons showed a reduction in firing rate for repeated stimuli, that is, 'repetition suppression' (likelihood ratio test, $P = 6.67 \times 10^{-9}$, $\chi^2(33.63,1)$) (Supplementary Fig. 6). Furthermore, among the neurons displaying significant repetition suppression, the relative decrease in firing rate for repeat presentations was greater at more anterior locations (Fig. 2). Interestingly, the recognition memory and spatial representations were somewhat independent of each other; at more anterior recording locations, grid cells were more likely to also exhibit a memory response (likelihood ratio test, $P = 0.0034$, $\chi^2(8.55,1)$), and the magnitude of firing-rate reduction was in line with the rest of the population of cells. However, there was a gradual decline in the percentage of grid cells at more anterior locations, and no grid cells were found in any of 3 separate penetrations in front of the posterior 50% of the EC, suggesting a functional border between posterior and anterior EC. Because most of our recordings were in the

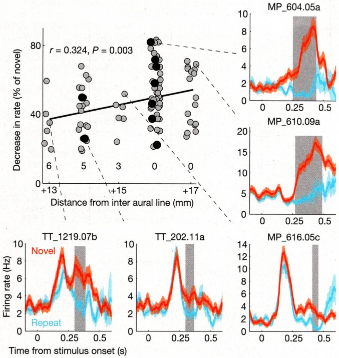


Figure 2 | Recognition memory and conjunctive grid-memory cells. The strength of the memory response (percentage decrease in rate for repeated stimuli) increased in more anterior recording locations ($n = 85$ cells with a significant recognition memory response). Grid cells (large black circles) were more likely to show a recognition memory response at more anterior locations (the number of grid cells without a significant memory response is given below the cluster of points at each location). Neuronal responses to novel and repeat presentations are shown for some representative neurons (right and bottom panels; mean ± s.e.m.). Grey shading indicates a time region of significant decrease in firing rate for repeat presentations ($P < 0.025$; Supplementary Methods). The monkeys' names and unit numbers are indicated above the graphs.

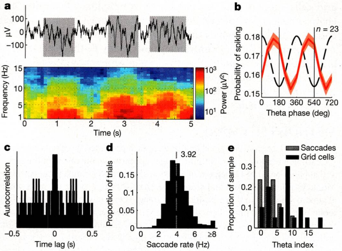


Figure 3 | Theta bouts and theta modulation of grid cells. a, Top, 5-s raw LFP trace with theta (3–12-Hz) bouts indicated in grey. Bottom, spectrogram showing the power (μV^2) of the top LFP trace. **b**, Grid cells were phase locked to LFP theta (60-degree bins, red curve, mean ± s.e.m.; dashed curve, prototype theta oscillation, two cycles are shown). **c**, Autocorrelation showing theta modulation of the spike train of a grid cell (theta index = 11.3 at 4.2 Hz). **d**, Histogram of the saccade rate during image viewings in 19 sessions (4.16 ± 1.49 Hz, mean ± s.d.; dashed grey line indicates the median, 3.92). **e**, Saccade times ($n = 17$ sessions) resulted in lower theta indices than the grid-cell spike trains ($n = 20$) ($P < 0.01$, Wilcoxon rank-sum test).

posterior EC, an important target of future studies will be to further characterize responses throughout the anterior EC.

EC layers were classified using current-source density estimates (see Supplementary Methods). Grid cells were found at an approximately equal frequency across both superficial and deep layers (14 out of 125 in superficial layers and 9 out of 68 in deep layers), suggesting that grid cells have a role in processing both input to and output from the hippocampus. Because many grid cells in layers deeper than layer II are modulated by head direction in rodents, it is possible that recording with the monkey's head positioned in variable directions would increase these percentages. In addition, the physical restriction of using dorsal to ventral penetrations precludes higher sampling of the thin layer II that possesses the largest percentage of grid cells in rodents.

As theta-band modulation is critical to certain models that describe the generation of grid cells[23], we next examined theta-band oscillatory activity in individual EC grid cells and in the simultaneously recorded local-field potential (LFP) on an adjacent electrode. We found that the EC exhibits intermittent bouts of theta oscillations similar to the theta bouts described previously in the primate hippocampus (Fig. 3a)[24,25]. These bouts, detected during blocks of image presentations, had a mean duration of 1.09 ± 0.86 s and a mean inter-bout interval of 0.95 ± 0.94 s (mean ± s.d.). We also found that across the recording session, the grid cells were phase-locked to near the trough of the LFP theta (Fig. 3b) and the spike trains of 13 out of 23 grid cells (57%) were theta-modulated (Fig. 3c), consistent with findings in rodents[26,27]. Theta modulation and phase locking were not limited to grid cells but were observed in the population of EC and HPC neurons as a whole (Supplementary Fig. 7c, d). During image viewing, saccadic eye movements were made at a median rate of 3.92 Hz (Fig. 3d), but the precise timing of saccades was not theta modulated across the session, suggesting that the theta modulation of grid cells was not due to rhythmicity in eye movements (Fig. 3e). Firing-rate maps for bout periods were correlated with non-bout periods ($P < 0.01$, Wilcoxon signed-rank test) and we did not find a significant difference in the grid pattern between bouts and non-bouts (Supplementary Fig. 8). Future experiments involving disruption of theta would aid in understanding the relationship between primate theta and spatial firing patterns[9,10]. Importantly, the present data demonstrate theta-band modulation among EC grid cells, even in a species with non-continuous theta. This information is critical for informing computational models and increasing our understanding of the generation of these spatial representations[23,28].

Taken together, these data provide evidence for grid cells in the primate entorhinal cortex. These spatial responses are similar in many respects to grid cells that have been described previously in the rat and bat during locomotion, but were identified with a distinct method of sampling the environment, that is, visual exploration through eye movements. These results suggest that spatial representations in primates can arise during visual exploration at a distance, without requiring an actual visit to that place. Accordingly, these data suggest that theories of spatial representation in the context of navigation could be applied to visual exploration. Because these experiments were performed with the monkey's head and the visual stimulus in a given, fixed position, we are not able to identify whether these grid cells represent allocentric or egocentric spatial locations. Notably, although grid-cell spiking was theta modulated, we found significant evidence for grid cells in the absence of continuous theta-band oscillations in the LFP. These data suggest that current models of grid cells based on interactions of oscillations will need to be adapted to account for both intermittent theta and exploration through saccadic movements. One possibility is that saccades produce a theta-band phase shift in the LFP. This could modulate the theta-band frequency in a way that might be comparable to the modulation in frequency that occurs through movement velocity in rodents. It is also likely that an optimal model will include aspects of current oscillatory interaction and network attractor models[29]. These results provide a potential challenge to the view that grid cells support path integration by combining self motion and environment cues[4]; however, it will be important to determine whether corollary discharge signals that enable the monitoring of eye movement commands might contribute significantly to the generation of grid fields[30]. It is possible that the spatial representations identified here reflect an integration of environmental context and the magnitude of eye movements, but this will need to be carefully examined in future experiments.

METHODS SUMMARY

In each VPLT session, 200 novel images were each presented twice on a computer monitor in a random order, with the images appearing at the centre of the screen and subtending 11×11 degrees of visual angle from the monkey's perspective. Monkeys were head fixed and seated in a chair with the centre of the monitor aligned to their neutral eye position. An image was removed after the monkey looked outside of the image bounds, or after 5 s, whichever was shorter. In a second version of the task that was used to record from 7 EC units, images covered the entire viewable region of the monitor (33×25 degrees of visual angle). In this version, each of 36 novel scenes was shown twice and a total of 10 s of visual exploration was required for each presentation. Gaze location was recorded with an infrared eye-tracking system (ISCAN). We recorded spikes (250–8,000 Hz) and LFPs (0.7–170 Hz) from the EC with a laminar electrode array mounted on a tungsten microelectrode (12-site, 150-μm spacing; FHC). Rate maps were computed with a Gaussian smoothing procedure[27]. Gridness scores and border scores were calculated using standard equations and significance was evaluated with a standard shuffling procedure[5,27]. Theta-band bouts in the LFP were detected by statistically analysing the power spectrum over time[25]. Theta-band modulation of grid cells was analysed with a standard autocorrelogram power spectrum metric, the theta index[27]. All experiments were carried out in accordance with protocols approved by the Emory University Institutional Animal Care and Use Committee.

Received 3 April; accepted 14 September 2012.

Published online 28 October; corrected online 28 November 2012 (see full-text HTML version for details).

1. Ekstrom, A. D. *et al.* Cellular networks underlying human spatial navigation. *Nature* **425,** 184–188 (2003).
2. Moser, E. I., Kropff, E. & Moser, M.-B. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* **31,** 69–89 (2008).
3. Yartsev, M. M., Witter, M. P. & Ulanovsky, N. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature* **479,** 103–107 (2011).
4. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436,** 801–806 (2005).
5. Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B. & Moser, E. I. Representation of geometric borders in the entorhinal cortex. *Science* **322,** 1865–1868 (2008).
6. Jutras, M. J. & Buffalo, E. A. Recognition memory signals in the macaque hippocampus. *Proc. Natl Acad. Sci. USA* **107,** 401–406 (2010).
7. Jutras, M. J., Fries, P. & Buffalo, E. A. Gamma-band synchronization in the macaque hippocampus and memory formation. *J. Neurosci.* **29,** 12521–12531 (2009).
8. Hafting, T., Fyhn, M., Bonnevie, T., Moser, M.-B. & Moser, E. I. Hippocampus-independent phase precession in entorhinal grid cells. *Nature* **453,** 1248–1252 (2008).
9. Koenig, J., Linder, A. N., Leutgeb, J. K. & Leutgeb, S. The spatial periodicity of grid cells is not sustained during reduced theta oscillations. *Science* **332,** 592–595 (2011).
10. Brandon, M. P. *et al.* Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science* **332,** 595–599 (2011).
11. Rolls, E. T. Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus* **9,** 467–480 (1999).
12. Tamura, R., Ono, T., Fukuda, M. & Nakamura, K. Spatial responsiveness of monkey hippocampal neurons to various visual and auditory stimuli. *Hippocampus* **2,** 307–322 (1992).
13. Ono, T., Nakamura, K., Nishijo, H. & Eifuku, S. Monkey hippocampal neurons related to spatial and nonspatial functions. *J. Neurophysiol.* **70,** 1516–1529 (1993).
14. Robertson, R. G., Rolls, E. T., Georges-François, P. & Panzeri, S. Head direction cells in the primate pre-subiculum. *Hippocampus* **9,** 206–219 (1999).
15. Suzuki, W. A., Miller, E. K. & Desimone, R. Object and place memory in the macaque entorhinal cortex. *J. Neurophysiol.* **78,** 1062–1081 (1997).
16. Doeller, C. F., Barry, C. & Burgess, N. Evidence for grid cells in a human memory network. *Nature* **463,** 657–661 (2010).
17. Witter, M. P., Wouterlood, F. G., Naber, P. A. & Van Haeften, T. Anatomical organization of the parahippocampal-hippocampal network. *Ann. NY Acad. Sci.* **911,** 1–24 (2000).
18. Insausti, R. & Amaral, D. G. Entorhinal cortex of the monkey: IV. Topographical and laminar organization of cortical afferents. *J. Comp. Neurol.* **509,** 608–641 (2008).
19. Witter, M. P. & Amaral, D. G. Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. *J. Comp. Neurol.* **307,** 437–459 (1991).
20. Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312,** 758–762 (2006).
21. Kjelstrup, K. B. *et al.* Finite scale of spatial representation in the hippocampus. *Science* **321,** 140–143 (2008).
22. Canto, C. B., Wouterlood, F. G. & Witter, M. P. What does the anatomical organization of the entorhinal cortex tell us? *Neural Plast.* **2008,** 381243 (2008).
23. Burgess, N., Barry, C. & Keefe, J. O. An oscillatory interference model of grid cell firing. *Hippocampus* **17,** 801–812 (2007).
24. Stewart, M. & Fox, S. E. Hippocampal theta activity in monkeys. *Brain Res.* **538,** 59–63 (1991).
25. Ekstrom, A. D. *et al.* Human hippocampal theta activity during virtual navigation. *Hippocampus* **15,** 881–889 (2005).
26. Mizuseki, K., Sirota, A., Pastalkova, E. & Buzsáki, G. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* **64,** 267–280 (2009).
27. Langston, R. F. *et al.* Development of the spatial representation system in the rat. *Science* **328,** 1576–1580 (2010).
28. Burgess, N. & O'Keefe, J. Models of place and grid cell firing and theta rhythmicity. *Curr. Opin. Neurobiol.* **21,** 734–744 (2011).
29. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the "cognitive map". *Nature Rev. Neurosci.* **7,** 663–678 (2006).
30. Sommer, M. A. & Wurtz, R. H. A pathway in primate brain for internal monitoring of movements. *Science* **296,** 1480–1482 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Potter, C. Erickson, J. Manns, M. Meister and K. Dunne for comments on the manuscript, and M. Tompkins and D. Solyst for assistance with experiments. This project was funded by the National Institute of Mental Health, R01MH093807 (E.A.B.), R01MH080007 (E.A.B.), MH082559 (M.J.J.), the National Center for Research Resources P51RR165, and is currently supported by the Office of Research Infrastructure Programs/OD P51OD11132. N.J.K. was supported by the NSF IGERT program (DGE-0333411).

Author Contributions E.A.B. and N.J.K. designed the research, N.J.K. collected the data from the entorhinal cortex, M.J.J. collected data from the hippocampus, N.J.K. performed the analyses, and N.J.K. and E.A.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A.B. (elizabeth.buffalo@emory.edu).

doi:10.1038/nature11463

Self-renewal of embryonic-stem-cell-derived progenitors by organ-matched mesenchyme

Julie B. Sneddon[1]*, Malgorzata Borowiak[1]* & Douglas A. Melton[1]

One goal of regenerative medicine, to use stem cells to replace cells lost by injury or disease, depends on producing an excess of the relevant cell for study or transplantation. To this end, the stepwise differentiation of stem cells into specialized derivatives has been successful for some cell types[1–3], but a major problem remains the inefficient conversion of cells from one stage of differentiation to the next. If specialized cells are to be produced in large numbers it will be necessary to expand progenitor cells, without differentiation, at some steps of the process. Using the pancreatic lineage as a model for embryonic-stem-cell differentiation, we demonstrate that this is a solvable problem. Co-culture with organ-matched mesenchyme permits proliferation and self-renewal of progenitors, without differentiation, and enables an expansion of more than a million-fold for human endodermal cells with full retention of their developmental potential. This effect is specific both to the mesenchymal cell and to the progenitor being amplified. Progenitors that have been serially expanded on mesenchyme give rise to glucose-sensing, insulin-secreting cells when transplanted *in vivo*. Theoretically, the identification of stage-specific renewal signals can be incorporated into any scheme for the efficient production of large numbers of differentiated cells from stem cells and may therefore have wide application in regenerative biology.

Several *in vitro* protocols have been devised to direct differentiation of pluripotent cells into mature cells of interest. Most successful approaches promote the transition of cells through a series of intermediates designed to mimic normal development[1–3]. In the pancreas, this entails progressing from embryonic stem cells (ESCs) (marked by expression of octamer-binding protein 4 (Oct4; also known as Pou5f1)) to definitive endoderm (marked by expression of the transcription factor SRY-box containing gene 17 (Sox17)), then pancreatic progenitors (marked by expression of the transcription factor pancreatic and duodenal homeobox1 (Pdx1)), endocrine progenitors (marked by expression of the transcription factor neurogenin 3 (Ngn3)), and finally mature β-cells (which express insulin; Fig. 1a). So far, most attention has focused on the signals responsible for directing differentiation from one stage to the next. Here we focus on amplifying or renewing distinct progenitors at various steps along the pancreatic lineage.

As the microenvironment has an important role in regulating the balance between renewal and differentiation for many pluripotent cells[4,5], we chose to co-culture ESC-derived progenitors with mesenchymal or endothelial cells, both of which influence embryonic pancreatic development *in vivo*[6–8]. Primary mesenchymal cell lines were established from embryonic, neonatal and adult mouse pancreas, intestine, liver and spleen, and from human pancreas (Supplementary Table 1). From a total of 68 primary samples, 16 yielded spindle-shaped cells that could be passaged at least 10 times. This panel of mesenchymal cells was tested for the ability to promote self-renewal of two distinct and transient progenitor cells in the pancreatic lineage: definitive endoderm and endocrine progenitors (Fig. 1a).

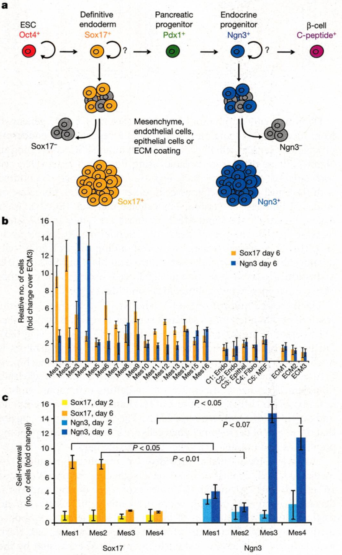


Figure 1 | Screen for signals that expand definitive endoderm and endocrine progenitors. a, Schema for directed differentiation of β-cells and their progenitors. **b**, Number of Sox17–GFP$^+$ cells or Ngn3–GFP$^+$ cells after co-culture with primary mesenchyme lines (Mes1 through to Mes16), control endothelial cell lines (C1, C2), an epithelial cell line (C3), a fibroblast cell line (C4), MEFs (C5) or various ECM surfaces (ECM1, ECM2 and ECM3) for 6 days. **c**, The number of cells (Sox17$^+$ and Ngn3$^+$) after 2 or 6 days of co-culture. *P* values were calculated using Student's *t*-test. Data represent the mean of two biological replicates ± s.d.

---

[1] Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts, 02138 USA.

*These authors contributed equally to this work.

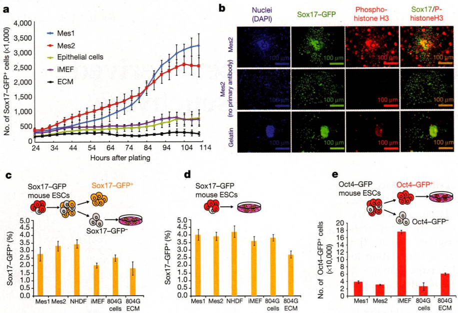


Figure 2 | Effects of mesenchyme are due to proliferation, not induction, and are specific to the responding cell type. **a**, Number of Sox17–GFP$^+$ cells over time during co-culture. **b**, Immunofluorescence staining of Sox17–GFP$^+$ cells after 48 h co-culture with Mes2 or gelatin. **c**, Number of Sox17–GFP$^+$ induced from sorted Sox17–GFP$^-$ cells by 6 days of co-culture. **d**, Number of Sox17–GFP$^+$ cells induced from mouse ESCs by 6 days of co-culture. **e**, Number of Oct4–GFP$^+$ cells after co-culture. iMEF, irradiated MEFs; NHDF, normal human dermal fibroblasts. 804G ECM, extracellular matrix from 804G cells. Data represent the mean of two biological replicates ± s.d.

Definitive endoderm (Sox17-positive (Sox17$^+$) cells)[9] generated from mouse ESCs containing a Sox17–GFP reporter and endocrine progenitors (Ngn3$^+$ cells)[10] generated from mouse ESCs containing an Ngn3–GFP reporter were isolated using fluorescence activated cell sorting (FACS) (Supplementary Figs 1 and 2) and then co-cultured with the panel of mesenchymal cells, other endothelial, fibroblast or epithelial lineages, or various extracellular matrices (ECMs). After 6 days in culture, Mes1 and Mes2 produced a 9.7-fold and 12.1-fold increase in the number of definitive endoderm cells, respectively (Fig. 1b). This effect is specific to the responding cell, as Ngn3$^+$ cells did not substantially expand when cultured on Mes1 or Mes2 (Fig. 1b, c). Instead, a marked expansion of Ngn3$^+$ cells occurs in the presence of two other mesenchymal lines, Mes3 and Mes4 (Fig. 1b).

The increased number of Sox17$^+$ and Ngn3$^+$ cells (Fig. 1c) is the result of proliferation, not a selective survival effect, as shown by live-cell imaging (Fig. 2a) and phospho-histone H3 staining (Fig. 2b). Furthermore, the mesenchyme-mediated renewal of progenitors is not the result of induction of Sox17 (Fig. 2c, d), preferential attachment to mesenchymal cells (Supplementary Fig. 3) or a decrease in cell death (Supplementary Fig. 4). In contrast to co-culture with mouse embryonic fibroblasts (MEFs), co-culture of Mes1 and Mes2 does not expand undifferentiated (Oct4$^+$) mouse ESCs (Fig. 2e). These data suggest that the effects of the mesenchymal cells are specific to the responding population and not due to a generic mitogenic signal.

The self-renewal signal provided by mesenchyme might be mediated by cell contact or by secreted signalling factors. We tested 16 growth factors, including those expressed during endoderm development and those with widespread mitogenic effects, but no single factor or combination of factors was sufficient to reproduce the magnitude of effect of co-culture (Supplementary Fig. 5a). Next, we attempted to block the mesenchyme-mediated expansion by the addition of a panel of 41 chemical inhibitors that cover a broad range of signalling pathways (Supplementary Fig. 5b). Some small molecules reduced the mesenchyme-mediated expansion by varying degrees, suggesting that there may be multiple pathways involved. Taken together, these data indicate that the expansion on mesenchyme is likely to be multifactorial. Future studies aimed at elucidating the mechanism behind mesenchyme-mediated expansion will delve further into the complexities of combinations of growth factors, extracellular matrix (ECM) proteins and chemical compounds.

Self-renewal without differentiation has perhaps been best studied for ESCs, for which self-renewal is defined as the ability of a cell to repeatedly divide without loss of identity or functional potential[11,12]. We tested whether long-term self-renewal of both mouse and human ESC-derived endoderm can be achieved *in vitro* by serial passage on Mes1 or Mes2. We observed a 3-million-fold and 6-million-fold expansion of mouse Sox17$^+$ cells on Mes1 and Mes2, respectively, after 7 passages (Fig. 3a), and a 65-million-fold expansion of human Sox17$^+$/FoxA2$^+$ cells on Mes2 after 9 passages (Fig. 3c; for data on mouse Sox17$^+$ cells that were successively sorted at each passage, see Supplementary Fig. 6). Global gene-expression analysis of mouse Sox17$^+$ cells expanded on Mes1 or Mes2 shows a very close concordance ($R^2 = 0.92$ and 0.96, respectively) between the average gene-expression level of all genes before and after 6 passages (Supplementary Fig. 7). The Sox17$^+$ endoderm expanded by mesenchyme continues to co-express markers of definitive endoderm (including forkhead box A2 (FoxA2) protein) and does not differentiate further, as indicated by no increase in the expression of pancreatic (Pdx1), intestinal (caudal type homeobox 2 (Cdx2))[13] or lung (SRY-box containing gene 2 (Sox2))[14] markers (Supplementary Fig. 8a, b).

To address further the question of whether cellular identity is preserved, we assessed the developmental potential of amplified Sox17$^+$ cells. Mouse ESC-derived endoderm cells, expanded on mesenchyme for six passages, showed no loss in their capacity for differentiation into pancreatic progenitors (marked by expression of Pdx1), endocrine progenitors (marked by expression of Ngn3) and β-like cells (marked by expression of C-peptide) (Supplementary Fig. 9). The efficiency of induction towards pancreatic lineages did not change between unpassaged (P0) and late passage mouse ESC (P6) or human (P8) ESC-derived endoderm as judged by the percentage of Pdx1$^+$, Ngn3$^+$, or C-peptide$^+$ cells at each stage (Fig. 3b and 3d). Culture with mesenchyme thus permits mouse and human endoderm self-renewal, defined as long-term expansion without alteration of the pattern of gene expression or developmental potential.

Finally, we subjected the expanded, human ESC-derived pancreatic cells to a stringent test: whether they can form insulin-expressing, glucose-responsive cells *in vivo*. The most efficient published protocols for *in vitro* differentiation of pluripotent cells to β-cells yield only a small percentage (typically 0–15%) of insulin-positive cells, and these cells do not secrete insulin in a glucose-responsive manner. Thus, to test physiologic potential, stem cells are differentiated *in vitro* to a progenitor stage and then implanted *in vivo* where they 'mature' to functional cells[1]. Human ESCs were differentiated *in vitro* to definitive endoderm and then expanded on mesenchyme for 3 to 7 passages

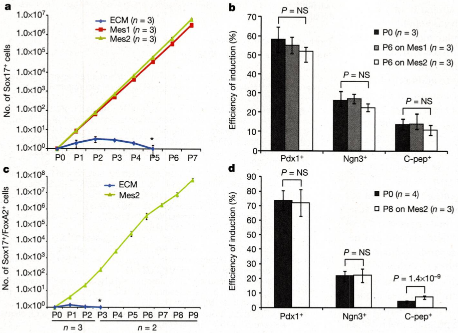


Figure 3 | Long-term expansion of differentiation-competent mouse and human ESC-derived endoderm in the presence of mesenchyme. **a**, Number of mouse Sox17–GFP$^+$ cells shown in relation to co-culture time. **b**, Efficiency of directed differentiation of unpassaged and passaged mouse Sox17–GFP$^+$ cells. **c**, Number of human Sox17$^+$ cells shown in relation to co-culture time. **d**, Efficiency of directed differentiation of unpassaged and passaged human Sox17$^+$ cells. Data represent mean of biological replicates ± s.d. *P* values are based on two-tailed Student's *t*-test. Px, passage number; 6 days (mouse) or 5 to 8 days (human) between passages. Asterisk denotes progressive loss of cells cultured on ECM alone.

(Fig. 4a). This expanded endoderm was then differentiated further *in vitro* to pancreatic progenitors and endocrine progenitors. Each cell type (expanded endoderm, pancreatic progenitors differentiated from expanded endoderm and endocrine progenitors differentiated from expanded endoderm, as well as unpassaged controls for each of the respective stages) was injected under the kidney capsule of SCID-Beige mice and allowed to mature *in vivo*. Before implantation, an aliquot of cells was fixed and stained to assess the state of differentiation. As expected, very few insulin (C-peptide)-expressing cells were detected at the endoderm and pancreatic progenitor stages *in vitro* (before transplantation; data not shown) and few (<5%) C-peptide$^+$ cells were detected at the endocrine progenitor stage (Supplementary Fig. 10).

All stages of human ESC-derived cells that were expanded by mesenchyme gave rise to Pdx1$^+$, C-peptide$^+$ (insulin-expressing) cells when transplanted *in vivo*. Immunofluorescence showed Pdx1, C-peptide and insulin co-expression in grafts of both unpassaged and passaged endoderm. Representative images for P7 passaged endoderm and its derivative P7 pancreatic progenitors are shown in Fig. 4b. In addition, human C-peptide was detected in the plasma of animals that received grafts of both unpassaged and passaged endoderm, pancreatic progenitors and endocrine progenitors (Supplementary Fig. 11). No C-peptide was detected in animals that received negative controls (in which PBS or mesenchyme alone were implanted). Human islets were also used as a positive control for engraftment, survival and function.

Most importantly, the implanted pancreatic progenitors from unpassaged and passaged endoderm secreted human C-peptide in a glucose-responsive manner (Fig. 4d). Animals that had been engrafted with pancreatic progenitors or controls were fasted for 16 h, then challenged by glucose injection. Negative controls included animals into which PBS or mesenchyme alone had been engrafted; human islets served as the positive control. The human islet controls show that this *in vivo* implantation assay has an inherent variability owing to difficulties in delivering the same number of cells to the kidney capsule,

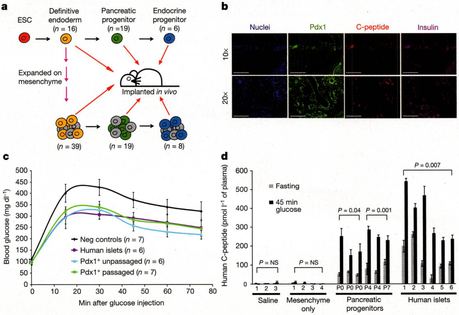


Figure 4 | Human ESC-derived cells expanded on mesenchyme give rise to insulin-expressing, glucose-responsive cells *in vivo*. **a**, Schematic depicting implantation of human ESC-derived progenitors. **b**, Immunofluorescence staining of human ESC-derived endoderm, passaged seven times on mesenchyme and engrafted for 3 months (top panel) or further differentiated to Pdx1$^+$ stage and then engrafted for 2 months (bottom panel). **c**, Glucose-tolerance test of animals implanted with PBS or mesenchyme only, human islets or Pdx1$^+$ pancreatic progenitors derived from unpassaged (P0), or passaged (P4 or P7) human endoderm. **d**, Fasting- and glucose-induced (45 min glucose) plasma human C-peptide levels. Pairs of bars represent two time points per animal; data represent mean of two technical replicates ± s.d.

as well as their engraftment and survival. The similarity of glucose-stimulated insulin secretion for the ESC-derived populations and human islet controls is notable, given that similar numbers of both cell types were implanted but the human islets have a much higher starting proportion of mature, insulin-expressing cells compared to the mixed population of pancreatic progenitors. In addition, glucose-tolerance tests revealed that compared to control animals, animals that had received Pdx1$^+$-stage pancreatic progenitors (either passaged or unpassaged) displayed a lower peak blood glucose, at levels similar to subjects that had received human islets (Fig. 4c). These *in vivo* implantation experiments provide evidence that mesenchyme-derived signals not only expand cells *in vitro* but give rise to cells that are physiologically relevant and functional in an *in vivo* context.

During embryogenesis, specification of progenitors is followed by amplification and further differentiation, and the balance between the two is probably responsible for determining the final organ size[15]. We show here that these two steps, renewal and differentiation, can be effectively uncoupled *in vitro*, enabling the separate control and manipulation of each step. This approach permits expansion of progenitors to an extent that may exceed that which occurs in normal *in vivo* development. Although we used the pancreatic lineage as a model, amplification of progenitors using organ-matched mesenchyme could be applicable for other tissue types and facilitate progress towards the goals of regenerative medicine.

METHODS SUMMARY

Progenitor–mesenchyme co-culture. Mouse or human ESCs were differentiated to either the definitive endoderm or endocrine progenitor stages, then cultured on top of mitotically inactivated, primary mesenchymal cells that had been derived from mouse or human tissue from various stages of development. As controls, progenitor cells were cultured on ECM surfaces or on mitotically inactivated control cell lines. The number of progenitors at the end of 6 days was assessed using immunofluorescence (in the case of human progenitors) and/or FACS analysis (when mouse ESC reporter lines were used). The following mouse ESC lines were used: Sox17-GFP[16,17], Ngn3-GFP[10] and Oct4–GFP (Oct4–GFP ESC lines were derived from Oct4–GFP mice from Jackson Laboratories)[18]. The human ESC line used was HUES8.

Growth factor and chemical screens. Recombinant growth factors were resuspended according to the manufacturer's instructions and used at a concentration of $20\,\mathrm{ng\,ml^{-1}}$and $50\,\mathrm{ng\,ml^{-1}}$. Chemicals were resuspended in DMSO (dimethylsulphoxide) at 10 mM concentration to prepare stock solutions. Human endoderm was co-cultured with Mes2 overnight, then treated with each compound at 1 μM and 10 μM final concentrations, with duplicates for both. After 6 days in the presence of compounds, cells were fixed and stained for Sox17 and FoxA2. Automated imaging and quantification were carried out using the ArrayScan (Cellomics), with at least 40 fields of view per well imaged in a 96-well plate.

Implantation of human ESC-derived progenitors *in vivo*. Human ESC-derived progenitors from multiple stages of differentiation were cultured alone (unpassaged) or co-cultured with Mes2 for 3 to 7 generations (passaged). Cells were then concentrated to a small volume corresponding to approximately 1.5 million cells in 50 μl of media for injection into the kidney capsule of SCID-Beige mice. Starting as early as 4 weeks post surgery, mice were administered glucose tolerance tests in which glucose was injected intraperitoneally ($3\,\mathrm{g\,kg^{-1}}$).

Full Methods and any associated references are available in the online version of the paper.

Received 8 November 2010; accepted 31 July 2012.

Published online 7 October 2012.

1. Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo*. *Nature Biotechnol.* **26,** 443–452 (2008).
2. Grigoriadis, A. E. *et al.* Directed differentiation of hematopoietic precursors and functional osteoclasts from human ES and iPS cells. *Blood* **115,** 2769–2776 (2010).
3. Perrier, A. L. *et al.* Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **101,** 12543–12548 (2004).
4. Moore, K. A. & Lemischka, I. R. Stem cells and their niches. *Science* **311,** 1880–1885 (2006).
5. Yamashita, Y. M. & Fuller, M. T. Asymmetric stem cell division and function of the niche in the *Drosophila* male germ line. *Int. J. Hematol.* **82,** 377–380 (2005).
6. Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from blood vessels. *Science* **294,** 564–567 (2001).
7. Golosow, N. & Grobstein, C. Epitheliomesenchymal interaction in pancreatic morphogenesis. *Dev. Biol.* **4,** 242–255 (1962).
8. Wessells, N. K. & Cohen, J. H. Early pancreas organogenesis: morphogenesis, tissue interactions, and mass effects. *Dev. Biol.* **15,** 237–270 (1967).
9. Kanai-Azuma, M. *et al.* Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129,** 2367–2379 (2002).
10. Lee, C. S., Perreault, N., Brestelli, J. E. & Kaestner, K. H. Neurogenin 3 is essential for the proper specification of gastric enteroendocrine cells and the maintenance of gastric epithelial cell identity. *Genes Dev.* **16,** 1488–1497 (2002).
11. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23,** 7150–7160 (2004).
12. Becker, A. J., Mc, C. E. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197,** 452–454 (1963).
13. James, R. & Kazenwadel, J. Homeobox gene expression in the intestinal epithelium of adult mice. *J. Biol. Chem.* **266,** 3246–3251 (1991).
14. Que, J. *et al.* Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* **134,** 2521–2531 (2007).
15. Thompson, D. A. W. *On Growth and Form* (Cambridge Univ. Press, 1917).
16. Borowiak, M. *et al.* Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell* **4,** 348–358 (2009).
17. Kim, I., Saunders, T. L. & Morrison, S. J. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* **130,** 470–483 (2007).
18. Lengner, C. J. *et al.* Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell* **1,** 403–415 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Morrison for providing the Sox17–GFP reporter mouse ESC line and K. Kaestner for the Ngn3–GFP knock-in mouse ESCs. We also thank J. LaVecchio, G. Buruzula and B. Tilton for support for cell sorting, A. Kweudjeu for help with gene-expression experiments and human ESC culture, and C. Xie, C. Balatbat and K. Koszka for technical assistance. We are grateful to A. Tward and D. Cohen for critical reading of the manuscript. We thank J. Annes for assistance in obtaining human tissue samples and acknowledge the use of human tissues provided by the National Disease Research Interchange (NDRI), with support from National Institutes of Health grants 5 U42 RR006042-20 and K08 DK084206. J.B.S. is supported by the Howard Hughes Medical Institute. M.B. was supported by a grant from The Leona M. and Harry B. Helmsley Charitable Trust. D.A.M. is an investigator of the Howard Hughes Medical Institute.

Author Contributions J.B.S., M.B., and D.A.M. conceived and designed the research. J.B.S. and M.B. carried out the experiments, and J.B.S., M.B. and D.A.M. analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.M. (dmelton@harvard.edu).

doi:10.1038/nature11583

Impaired intrinsic immunity to HSV-1 in human iPSC-derived TLR3-deficient CNS cells

Fabien G. Lafaille[1,2]*, Itai M. Pessach[3,4]*, Shen-Ying Zhang[5,6]*, Michael J. Ciancanelli[5], Melina Herman[5,6], Avinash Abhyankar[5], Shui-Wang Ying[7], Sotirios Keros[8], Peter A. Goldstein[7], Gustavo Mostoslavsky[9], Jose Ordovas-Montanes[3], Emmanuelle Jouanguy[5,6], Sabine Plancoulaine[6], Edmund Tu[1,2], Yechiel Elkabetz[10], Saleh Al-Muhsen[11], Marc Tardieu[12], Thorsten M. Schlaeger[13], George Q. Daley[13], Laurent Abel[5,6], Jean-Laurent Casanova[5,6,14], Lorenz Studer[1,2] & Luigi D. Notarangelo[3,15]

In the course of primary infection with herpes simplex virus 1 (HSV-1), children with inborn errors of toll-like receptor 3 (TLR3) immunity are prone to HSV-1 encephalitis (HSE)[1-3]. We tested the hypothesis that the pathogenesis of HSE involves non-haematopoietic CNS-resident cells. We derived induced pluripotent stem cells (iPSCs) from the dermal fibroblasts of TLR3- and UNC-93B-deficient patients and from controls. These iPSCs were differentiated into highly purified populations of neural stem cells (NSCs), neurons, astrocytes and oligodendrocytes. The induction of interferon-β (IFN-β) and/or IFN-λ1 in response to stimulation by the dsRNA analogue polyinosinic:polycytidylic acid (poly(I:C)) was dependent on TLR3 and UNC-93B in all cells tested. However, the induction of IFN-β and IFN-λ1 in response to HSV-1 infection was impaired selectively in UNC-93B-deficient neurons and oligodendrocytes. These cells were also much more susceptible to HSV-1 infection than control cells, whereas UNC-93B-deficient NSCs and astrocytes were not. TLR3-deficient neurons were also found to be susceptible to HSV-1 infection. The rescue of UNC-93B- and TLR3-deficient cells with the corresponding wild-type allele showed that the genetic defect was the cause of the poly(I:C) and HSV-1 phenotypes. The viral infection phenotype was rescued further by treatment with exogenous IFN-α or IFN-β (IFN-α/β) but not IFN-λ1. Thus, impaired TLR3- and UNC-93B-dependent IFN-α/β intrinsic immunity to HSV-1 in the CNS, in neurons and oligodendrocytes in particular, may underlie the pathogenesis of HSE in children with TLR3-pathway deficiencies.

Childhood HSE is a rare, life-threatening, central nervous system (CNS)-restricted complication of primary infection with HSV-1, an almost ubiquitous virus that is typically innocuous[4]. Children with HSE are not unusually susceptible to other infectious agents, including viruses, or even to HSV-1-related diseases affecting sites other than the CNS[4,5]. HSV-1 reaches the CNS from the oral and nasal epithelium, via the cranial nerves[4]. We identified autosomal recessive UNC-93B deficiency as the first genetic aetiology of childhood HSE[1]. UNC-93B is required for TLR3, TLR7, TLR8 and TLR9 responses[1,6]. We then identified autosomal-recessive or autosomal-dominant deficiencies of TLR3 (refs 2 and 3), TRAF3 (ref. 7), TRIF[8] and TBK1 (ref. 9), revealing that childhood HSE can be due to the impairment of TLR3 immunity. HSV-1 produces double-stranded (dsRNA) during its replication[10,11] and the dsRNA-sensing TLR3 is expressed and functional in non-haematopoietic (neurons, astrocytes, oligodendrocytes) and haematopoietic (microglia) CNS-resident cells, which produce IFN-β and IFN-λ in response to TLR3 stimulation[12-15] and can be infected with HSV-1 *in vitro*[13,16-19]. We therefore surmised that the pathogenesis of HSE in patients with TLR3-pathway deficiencies may involve impaired TLR3-dependent IFN-α/β and/or IFN-λ immunity to HSV-1 in the CNS.

We tested this hypothesis by generating induced pluripotent stem cells (iPSCs) from control, UNC-93B- and TLR3-deficient dermal fibroblasts (Supplementary Table 1). We first derived and fully characterized iPSC lines from a healthy child, from an HSE child with autosomal-recessive complete UNC-93B deficiency[1], and from a patient with autosomal-recessive complete TLR3 deficiency[2], by reprogramming primary dermal fibroblasts, as described previously[20] (Supplementary Fig. 1 and Methods). A robust stemness and pluripotency profile, karyotypic integrity and patient-specific origin of the iPSCs were confirmed (Supplementary Fig. 1). Whole-exome sequencing for one control, one TLR3-deficient and two UNC-93B-deficient iPSC lines revealed more than 99.9% concordance with the corresponding parental fibroblast lines, in terms of exonic genetic variability (Supplementary Table 2). No synonymous or non-synonymous germline and somatic rare variants of any of the 21 known TLR3-pathway genes were found in parental fibroblast and iPSC lines, respectively (Supplementary Table 3). We also used two additional, previously described healthy control iPSC lines[21,22] for subsequent experiments (Supplementary Table 1).

We next induced the differentiation of iPSCs into all major non-haematopoietic CNS-resident cells, including neural stem cells (NSCs), neurons, oligodendrocytes and astrocytes[23,24]. The selective derivation of each individual neural lineage required a multistep iPSC-differentiation approach combined with fluorescence-activated cell sorting (FACS)-mediated cell purification (Fig. 1a). Neural differentiation of UNC-93B-deficient iPSCs, TLR3-deficient iPSCs, control iPSCs and control human embryonic stem cells (hESCs) (H9 line) was induced by dual SMAD inhibition[21,24] (Fig. 1a and Supplementary Fig. 2a). The resulting polarized columnar neuroepithelial structures expressed PLZF, ZO1 and PAX6, well-known markers of neural rosettes[23]. Mechanically

---

[1]Center for Stem Cell Biology, Sloan-Kettering Institute for Cancer Research, New York, New York 10065, USA. [2]Developmental Biology Program, Sloan-Kettering Institute for Cancer Research, New York, New York 10065, USA. [3]Division of Immunology, Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]The Talpiot Medical Leadership Program, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel-Hashomer and the Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 52621, Israel. [5]St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University, New York, New York 10065, USA. [6]Laboratory of Human Genetics of Infectious Diseases, Institut National de la Santé et de la Recherche Médicale, University Paris Descartes, Necker Medical School, U980, Paris 75015, France. [7]C.V. Starr Laboratory for Molecular Neuropharmacology, Department of Anesthesiology, Weill Cornell Medical College, New York, New York 10065, USA. [8]Division of Pediatric Neurology, Department of Pediatrics, Weill Cornell Medical College, New York, New York 10065, USA. [9]Section of Gastroenterology, Department of Medicine and Center for Regenerative Medicine (CReM), Boston University School of Medicine, Boston, Massachusetts 02118, USA. [10]Laboratory for Pluripotent and Neural Stem Cell Biology, Department of Cell and Developmental Biology, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. [11]Prince Naif Center for Immunology Research, Department of Pediatrics, College of Medicine, King Saud University, Riyadh 11451, Saudi Arabia. [12]Department of Pediatric Neurology, Assistance Publique-Hôpitaux de Paris, Bicêtre Hospital, Kremlin-Bicêtre, 94275, France. [13]Division of Pediatric Hematology/Oncology, Children's Hospital and Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [14]Pediatric Hematology-Immunology Unit, Necker Hospital, Paris 75015, France. [15]The Manton Center for Orphan Disease Research, Children's Hospital, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

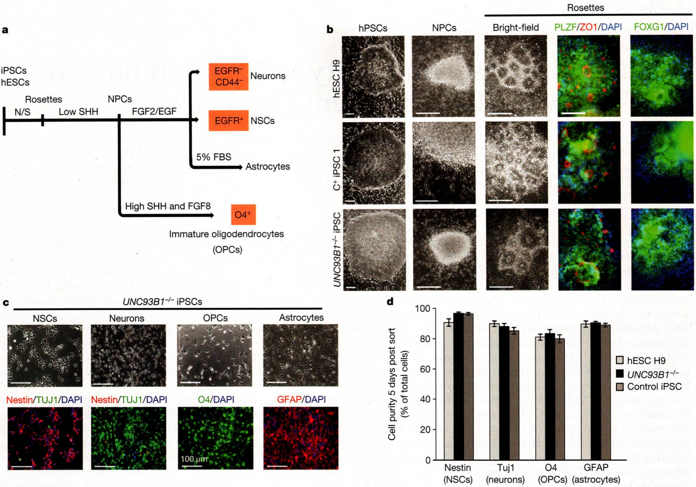

a
iPSCs
hESCs
Rosettes
N/S
Low SHH
NPCs
FGF2/EGF
EGFR⁻ CD44⁻
Neurons
EGFR⁺
NSCs
5% FBS
Astrocytes
High SHH and FGF8
O4⁺
Immature oligodendrocytes
(OPCs)
b
Rosettes
hPSCs
NPCs
Bright-field
PLZF/ZO1/DAPI
FOXG1/DAPI
hESC H9
C⁺ iPSC 1
UNC93B1⁻/⁻ iPSC
c
UNC93B1⁻/⁻ iPSCs
NSCs
Neurons
OPCs
Astrocytes
Nestin/TUJ1/DAPI
Nestin/TUJ1/DAPI
O4/DAPI
GFAP/DAPI
100 μm
d
Cell purity 5 days post sort
(% of total cells)
100
80
60
40
20
0
hESC H9
UNC93B1⁻/⁻
Control iPSC
Nestin
(NSCs)
Tuj1
(neurons)
O4
(OPCs)
GFAP
(astrocytes)

Figure 1 | Derivation and purification of CNS cells. a, Schematic diagram of the differentiation and purification protocols used. **b**, Representative phase-contrast images showing the morphology of the human pluripotent stem cells (hPSCs), neural rosettes and NPC clusters, from healthy control hESCs, healthy control iPSCs (C$^+$ (iPSC)) and *UNC93B1*$^{-/-}$ iPSCs. Immunocytochemistry analysis revealed the expression of rosette markers (PLZF and ZO1) and a forebrain marker FOXG1. **c**, Characterization of *UNC93B1*$^{-/-}$ iPSC-derived CNS cell types. Upper panels, phase-contrast images showing the characteristic morphology of each type of neural cell; lower panels, immunofluorescence analysis for markers of neural stem cells (nestin), neurons (TUJ1), oligodendrocyte progenitor cells (O4) and astrocytes (GFAP). **d**, Quantification of marker expression for each neural cell type derived from control hESCs, *UNC93B1*$^{-/-}$ iPSCs and control iPSCs (error bars, s.e.m.). Scale bars represent 100 μm (**b**), 50 μm (**c**; except for O4 staining). High SHH, high concentration of recombinant Sonic hedgehog (C25II (Cys25Ile-Ile)) at 20 ng ml^{-1}; low SHH, low concentration of recombinant Sonic hedgehog at 125 ng ml^{-1}; N/S, noggin and SB431542 (dual SMAD inhibition).

passaged rosettes retained expression of the forebrain marker FOXG1 (Fig. 1b) and yielded neural precursor cells (NPCs) in the presence of FGF2 and EGF (Fig. 1b and Supplementary Fig. 2a). Through differential growth-factor treatment and the use of cell-type-specific surface markers in NPC-stage cells, we were able to isolate highly enriched populations of NSCs, neurons, astrocytes or oligodendrocyte lineage cells (Fig. 1c and Supplementary Fig. 2b, c).

The purified hESC- and iPSC-derived neuronal populations expressed additional lineage-specific markers (Supplementary Fig. 3a) and showed the electrophysiological properties of functional neurons (Supplementary Fig. 3b–e). The identity of the purified glial fibrillary acidic protein (GFAP)-expressing and O4-antigen-expressing glial cell populations was confirmed by global gene expression profiling (Supplementary Fig. 4a, b). Immunocytochemistry for stage-specific markers was used to identify the purified O4 'oligodendrocytes' used throughout this study as a mostly immature (pre-myelinating) oligodendrocyte population (Supplementary Fig. 4c–e, Supplementary Table 4). Quantitative analysis showed that our preparations of NSCs, neurons, astrocytes and oligodendrocytes were highly pure (Fig. 1d), and similar gene-expression profiles were obtained for neurons and astrocytes derived from disease-specific and control cell lines (Supplementary Fig. 5). Our *in vitro* CNS cell-differentiation system therefore constitutes a reliable platform for the comparative assessment of CNS cell-specific antiviral immunity.

TLR3 expression has been documented in neurons derived *in vitro* from a human teratocarcinoma cell line[13], and in primary cells, either in human brain tissues *in situ* (neurons[25]) or isolated from human brain *ex vivo* (oligodendrocytes and astrocytes[12,14,26]), but not in human NSCs (self-renewing, multipotent cells responsible for generating neurons, astrocytes and oligodendrocytes in the CNS)[27]. We detected messenger RNA for key genes of the TLR3-responsive pathway, including *TLR3* and *UNC93B1*, in all four cell types tested (Supplementary Fig. 6a-d), whether differentiated from iPSCs or hESCs. We also detected mRNAs for genes encoding key molecules in the IFN-responsive pathways, including the receptors for IFN-α/β and IFN-λ, in these cells (Supplementary Fig. 6e–h). Levels of mRNA for the TLR3- and IFN-responsive pathway genes tested were similar, for each CNS cell type, between cells differentiated from control iPSCs, control hESCs and UNC-93B-deficient iPSCs, with the expected exception of *UNC93B1*, for which mRNA levels were lower in UNC-93B-deficient cells (Fig. 2a and Supplementary Fig. 6a–d).

We then studied TLR3 responses in the same cells. IFN-λ1 and IFN-β were induced in a time-dependent manner, by stimulation with the non-specific TLR3 agonist poly(I:C), a dsRNA analogue, in NSCs,

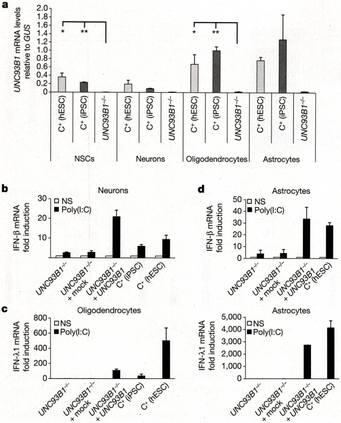


Figure 2 | UNC-93B-dependent IFN responses to TLR3 in neurons and glial cells. **a**, *UNC93B1* mRNA levels in CNS cells differentiated from hESCs from a healthy control (C^+ (hESC)) and iPSCs from a healthy control (C^+ (iPSC)), and an UNC-93B-deficient patient (*UNC93B1*$^{-/-}$). NS, not stimulated. **b**, **c**, IFN-β (**b**) or IFN-λ1 (**c**) mRNA induction, after 6 h of poly(I:C) stimulation, in neurons (**b**) or oligodendrocytes (**c**) differentiated from one healthy control hESC line, one healthy control iPSC line, and in *UNC93B1*$^{-/-}$ neurons (**b**) or oligodendrocytes (**c**), without lentiviral infection, or after infection with a lentivirus containing human wild-type *UNC93B1* (*UNC93B1*$^{-/-}$ + *UNC93B1*) or a mock construct (*UNC93B1*$^{-/-}$ + mock). **d**, IFN-β (upper panel) or IFN-λ1 (lower panel) mRNA induction, after 4 h (upper panel) or 6 h (lower panel) of poly(I:C) stimulation, in astrocytes differentiated from hESCs from a healthy control, in *UNC93B1*$^{-/-}$ astrocytes, without lentiviral infection, or after infection with a lentivirus containing human wild-type *UNC93B1* (*UNC93B1*$^{-/-}$ + *UNC93B1*) or a mock construct (*UNC93B1*$^{-/-}$ + mock). Mean values ± s.e.m. were calculated from three (**a**) or two (**b**–**d**) independent experiments, each carried out in duplicate. Analysis of variance (ANOVA) was carried out for *UNC93B1* mRNA expression levels shown in **a**. When significant, Dunnett's *t*-tests were performed for two-by-two comparisons. $^*P < 0.05$, $^{**}P < 0.01$ (comparisons between each control line and the patient line, for each cell type).

neurons, oligodendrocytes and astrocytes differentiated from healthy control iPSCs or hESCs, but not in UNC-93B-deficient iPSC-differentiated CNS cells (Fig. 2b–d and Supplementary Fig. 6i–l). The induction of NF-κB1, a key IFN-inducing molecule, and that of MX1, a key IFN-inducible molecule, were both impaired in UNC-93B-deficient oligodendrocytes after poly(I:C) stimulation (Supplementary Fig. 6m, n). The impaired response to poly(I:C) in UNC-93B-deficient CNS cells was consistent with our previous data for UNC-93B-deficient fibroblasts, pointing to a TLR3-dependent response to dsRNA in these cell types[1]. Moreover, impaired poly(I:C) responsiveness in UNC-93B-deficient neurons, oligodendrocytes and astrocytes was rescued by transient expression of the human *UNC93B1* gene (Fig. 2b–d). Thus, the UNC-93B-dependent TLR3 pathway is functional in control human iPSC-derived NSCs, neurons, astrocytes and oligodendrocytes, at least for the induction of antiviral IFN-λ1 and IFN-β in response to poly(I:C).

We thus set out to compare the response to HSV-1 in UNC-93B-deficient and control iPSC or hESC-derived CNS cells after infection with HSV-1 and HSV-1–GFP (green fluorescent protein)[28]. Human NSCs and astrocytes seemed to be more susceptible to HSV-1 infection than neurons and oligodendrocytes, as massive HSV-1–GFP replication was observed earlier (Supplementary Fig. 7 and data not shown). UNC-93B-deficient NSCs and astrocytes derived from two different iPSC lines also showed high levels of HSV-1-GFP replication, similar to those observed in the corresponding cell types derived from two control iPSC lines and the control hESC line (Fig. 3a–c and Supplementary Figs 8a–c and 9a–c). Treatment of UNC-93B-deficient and control NSCs and astrocytes with recombinant IFN-α2B or IFN-β, but not IFN-λ1, decreased HSV-1–GFP replication levels (Supplementary Figs 8a, c–f and 9a, c–g). Moreover, treatment with poly(I:C) decreased HSV-1–GFP replication levels in control, but not in UNC-93B-deficient astrocytes (Supplementary Figs 9f and 10a–d). By contrast, treatment with agonists of TLR9 (CpG-A or CpG-C) did not have such an effect (Supplementary Fig. 10a–d).

When UNC-93B-deficient neurons from the two UNC-93B-deficient iPSC lines were infected with HSV-1–GFP, HSV-1–GFP replication was faster, reaching higher levels than in neurons differentiated from four control iPSC lines or one hESC line (Fig. 3a, d, e and Supplementary Fig. 11a). The treatment of UNC-93B-deficient neurons with IFN-α2B or IFN-β, but not IFN-λ1, rescued this phenotype (Supplementary Fig. 11a). Similar results were obtained with TLR3-deficient neurons that had been differentiated from TLR3-deficient iPSCs[3,20] (Fig. 3e and Supplementary Fig. 11b). The phenotype of enhanced HSV-1 replication in UNC-93B- and TLR3-deficient neurons was rescued by expression of the wild-type human *UNC93B1* and *TLR3* genes, respectively (Fig. 3f and Supplementary Fig. 11b). Finally, higher levels and faster replication of HSV-1–GFP were also observed in UNC-93B-deficient oligodendrocytes than in oligodendrocytes differentiated from control iPSCs or hESCs, and this phenotype was rescued by treatment with IFN-α2B or IFN-β, but not IFN-λ1 (Fig. 3a, g and Supplementary Fig. 11c, d). This is consistent with our previous finding of high susceptibility to HSV-1 and VSV in fibroblasts with TLR3-pathway deficiencies, associated with an impairment of the TLR3-dependent induction of IFN-β and IFN-λ, which can be rescued more effectively by exogenous IFN-α/β than by IFN-λ1 (refs 1–3).

We studied further the production of IFNs and inflammatory cytokines in UNC-93B-deficient and control CNS cells after infection with HSV-1. UNC-93B-deficient neurons, astrocytes and oligodendrocytes produced normal amounts of interleukin-6 (IL-6), as shown by comparison with control iPSC- or hESC-differentiated CNS cell types (Supplementary Fig. 12a–c). UNC-93B-deficient NSCs and astrocytes seemed to produce detectable but lower levels of IFN-λ1 than the control cells tested (Supplementary Fig. 12d, e), suggesting that UNC-93B-independent, TLR3-independent partially compensatory pathways may be involved in triggering IFN responses to HSV-1 in human NSCs and astrocytes. The induction of IFN-β or IFN-λ1 was readily observed in control iPSC- or hESC-differentiated neurons or oligodendrocytes, but was greatly impaired in UNC-93B-deficient neurons and oligodendrocytes, respectively (Supplementary Fig. 12f, g). The induction of MX1 was also impaired in UNC-93B-deficient oligodendrocytes after HSV-1 infection (Supplementary Fig. 12h). Thus, neurons and oligodendrocytes lacking UNC-93B may be highly vulnerable to HSV-1 infection because of an impairment of the cell-autonomous IFN-α/β or IFN-λ immunity.

Our findings suggest that neurons and oligodendrocytes provide strong, intrinsic protective anti-HSV-1 immunity in the CNS, through an intact TLR3 pathway. Although NSCs and astrocytes lacking UNC-93B are not more susceptible to HSV-1 infection than control cells *in vitro*, they may have a role in protective anti-HSV-1 immunity in the CNS *in vivo*. Indeed, mouse astrocytes rely on TLR3 to control HSV-2 (ref. 29). Haematopoietic cell-derived microglial cells, which also

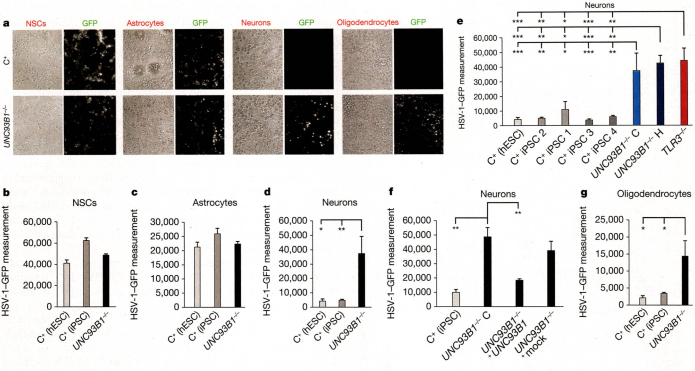


Figure 3 | High HSV-1 susceptibility in UNC-93B-deficient neurons and oligodendrocytes. Infection with HSV-1–GFP for 24 h, at a multiplicity of infection of 1, was carried out. **a**, GFP expression in CNS cells differentiated from *UNC93B1*$^{-/-}$ iPSCs or from hESCs from a healthy control (C$^+$) was visualized using fluorescence microscopy. Phase-contrast photomicrographs from the same view are also shown. **b–d** GFP expression in NSCs (**b**), astrocytes (**c**) and neurons (**d**) differentiated from *UNC93B1*$^{-/-}$ iPSCs, from hESCs from a healthy control (C$^+$ (hESC)) and one to two iPSC lines each from up to three healthy controls (C$^+$ (iPSC)), was assessed with a fluorescence plate reader. The difference in GFP intensity between HSV-1–GFP-infected cells and uninfected cells is shown. **e**, GFP expression in neurons differentiated from two lines (C and H) of *UNC93B1*$^{-/-}$ iPSCs, one line of *TLR3*$^{-/-}$ iPSCs, a total of four iPSC lines from three healthy controls, one or two lines from each, or from one C$^+$ hESCs line. **f**, GFP expression, in neurons from one C$^+$ iPSC line, in *UNC93B1*$^{-/-}$ neurons, without lentiviral infection, or after infection with a lentivirus containing human wild-type *UNC93B1* (*UNC93B1*$^{-/-}$ + *UNC93B1*) or a mock construct (*UNC93B1*$^{-/-}$ + mock). **g**, GFP expression in oligodendrocytes differentiated from *UNC93B1*$^{-/-}$ iPSCs, from a C$^+$ hESC line and a C$^+$ iPSC line. Error bars, s.e.m., calculated from two to three experiments, from the C$^+$ hESC lines, the C$^+$ iPSC lines and the *UNC93B1*$^{-/-}$ lines, each studied in triplicate (**b–d**, **e** and **g**); or from one experiment carried out in triplicate, representative of two independent experiments (**f**). ANOVA was performed for the HSV-1–GFP levels shown in **b** to **g**. When significant, Dunnett's *t*-tests were performed for two-by-two comparisons. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (comparisons between each control line and each patient line).

express TLR3 and can be infected with HSV-1 (refs 12, 14, 17), may also contribute to HSE. CNS-intrinsic mechanisms are thus vital for the control of HSV-1 in the course of primary infection in childhood. These new findings add to our previous results that indicated that most leukocytes and keratinocytes from HSE patients with TLR3-pathway deficiencies respond normally to stimulation with poly(I:C) or HSV-1 (refs 2, 3 and 9), and are consistent with the CNS-restricted pattern of lesions during childhood HSE, with no disseminated disease. Human non-haematopoietic cells may be the key to survival during the course of primary infection, extending the concepts of host defence beyond innate and adaptive haematopoietic immunity, to non-haematopoietic 'intrinsic' immunity[30].

METHODS SUMMARY

Human iPSC generation. Fibroblasts from patients or controls were transduced with the polycistronic stem cell cassette (STEMCCA) lentiviral vector and cultured in DMEM medium supplemented with 10% fetal calf serum (FCS), 2 mM L-glutamine, penicillin (50 U ml^{-1}) and streptomycin (50 μg ml^{-1}) as described previously[20]. After 72 h, cells were transferred onto irradiated mouse embryonic fibroblasts (iMEFs) and the medium was replaced with hESC medium. iPSC colonies with an ESC-like morphology were mechanically isolated 4 to 5 weeks after infection.

Neural differentiation. We adapted previously described protocols for neural differentiation[21,23]. Rosettes were collected on day 8 and cultured further to give clusters of proliferating NPCs. nestin$^+$/TUJ1$^-$ NSCs were maintained in N2 medium supplemented with epidermal growth factor (EGF) (20 ng ml^{-1}) and fibroblast growth factor 2 (FGF2) (20 ng ml^{-1}) (R&D Systems) and B-27 supplement (1:100, Invitrogen) and sorted for EGF-receptor (EGFR) expression. Nestin$^-$/TUJ1$^+$ neurons were obtained by eliminating EGFR$^+$ NSCs and CD44$^+$ non-neural cells from NPCs. For the generation of astrocytes, NPCs were allowed to proliferate in the presence of EGF and FGF2 for 40 to 60 days and were then cultured in medium containing 5% FBS for an additional 15 to 20 days. Immature oligodendrocytes were obtained from NPCs following treatment with SonicC25II (125 ng ml^{-1}), FGF8 (100 ng ml^{-1}; R&D Systems), brain-derived neurotrophic factor (BDNF) (20 ng ml^{-1}) and ascorbic acid (0.2 mM) for 50 to 70 days and sorted for O4.

TLR3 agonists and viral stimulation. A synthetic analogue of dsRNA, (polyinosinic:polycytidylic acid (poly(I:C)); Amersham), a TLR3 agonist, was used at a concentration of 25 μg ml^{-1}, for 2, 4 or 6 h of stimulation. For HSV-1 infection, we used the KOS-1 strain of HSV-1 and a GFP-expressing HSV-1 (HSV-1–GFP[28]), at various multiplicities of infection (MOIs), to infect CNS cells. The induction of IFN-β, IFN-λ and IFN-responsive genes was assessed in these cells, by quantifying mRNA by quantitative polymerase chain reaction with reverse transcription (RT–qPCR).

Full Methods and any associated references are available in the online version of the paper.

Received 25 October 2010; accepted 12 September 2012.

Published online 28 October; corrected online 28 November 2012 (see full-text HTML version for details).

1. Casrouge, A. *et al.* Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* **314,** 308–312 (2006).
2. Zhang, S. Y. *et al.* TLR3 deficiency in patients with herpes simplex encephalitis. *Science* **317,** 1522–1527 (2007).
3. Guo, Y. *et al.* Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. *J. Exp. Med.* **208,** 2083–2098 (2011).
4. Whitley, R. J. Herpes simplex encephalitis: adolescents and adults. *Antiviral Res.* **71,** 141–148 (2006).
5. Abel, L. *et al.* Age-dependent Mendelian predisposition to herpes simplex virus type 1 encephalitis in childhood. *J. Pediatr.* **157,** 623–629 (2010).

6. Kim, Y. M., Brinkmann, M. M., Paquet, M. E. & Ploegh, H. L. UNC93B1 delivers nucleotide-sensing toll-like receptors to endolysosomes. *Nature* **452,** 234–238 (2008).
7. Pérez de Diego, R. *et al.* Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity* **33,** 400–411 (2010).
8. Sancho-Shimizu, V. *et al.* Herpes simplex encephalitis in children with autosomal recessive and dominant TRIF deficiency. *J. Clin. Invest.* **121,** 4889–4902 (2011).
9. Herman, M. *et al.* Heterozygous TBK1 mutations impair TLR3 immunity and underlie herpes simplex encephalitis of childhood. *J. Exp. Med.* **209,** 1567–1582 (2012).
10. Jacquemont, B. & Roizman, B. RNA synthesis in cells infected with herpes simplex virus. X. Properties of viral symmetric transcripts and of double-stranded RNA prepared from them. *J. Virol.* **15,** 707–713 (1975).
11. Weber, F., Wagner, V., Rasmussen, S. B., Hartmann, R. & Paludan, S. R. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J. Virol.* **80,** 5059–5064 (2006).
12. Bsibsi, M., Ravid, R., Gveric, D. & van Noort, J. M. Broad expression of Toll-like receptors in the human central nervous system. *J. Neuropathol. Exp. Neurol.* **61,** 1013–1021 (2002).
13. Préhaud, C., Megret, F., Lafage, M. & Lafon, M. Virus infection switches TLR-3-positive human neurons to become strong producers of beta interferon. *J. Virol.* **79,** 12893–12904 (2005).
14. Jack, C. S. *et al.* TLR signaling tailors innate immune responses in human microglia and astrocytes. *J. Immunol.* **175,** 4320–4330 (2005).
15. Zhou, L. *et al.* Activation of Toll-like receptor-3 induces interferon-lambda expression in human neuronal cells. *Neuroscience* **159,** 629–637 (2009).
16. Mitchell, B. M., Bloom, D. C., Cohrs, R. J., Gilden, D. H. & Kennedy, P. G. Herpes simplex virus-1 and varicella-zoster virus latency in ganglia. *J. Neurovirol.* **9,** 194–204 (2003).
17. Lokensgard, J. R. *et al.* Robust expression of TNF-alpha, IL-1beta, RANTES, and IP-10 by human microglial cells during nonproductive infection with herpes simplex virus. *J. Neurovirol.* **7,** 208–219 (2001).
18. Bello-Morales, R., Fedetz, M., Alcina, A., Tabares, E. & Lopez-Guerrero, J. A. High susceptibility of a human oligodendroglial cell line to herpes simplex type 1 infection. *J. Neurovirol.* **11,** 190–198 (2005).
19. Marques, C. P., Hu, S., Sheng, W. & Lokensgard, J. R. Microglial cells initiate vigorous yet non-protective immune responses during HSV-1 brain infection. *Virus Res.* **121,** 1–10 (2006).
20. Pessach, I. M. *et al.* Induced pluripotent stem cells: a novel frontier in the study of human primary immunodeficiencies. *J. Allergy Clin. Immunol.* **127,** 1400–1407 (2011).
21. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27,** 275–280 (2009).
22. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480,** 547–551 (2011).
23. Elkabetz, Y. *et al.* Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* **22,** 152–165 (2008).
24. Tabar, V. *et al.* Migration and differentiation of neural precursors derived from human embryonic stem cells in the rat brain. *Nature Biotechnol.* **23,** 601–606 (2005).
25. Jackson, A. C., Rossiter, J. P. & Lafon, M. Expression of Toll-like receptor 3 in the human cerebellar cortex in rabies, herpes simplex encephalitis, and other neurological diseases. *J. Neurovirol.* **12,** 229–234 (2006).
26. Farina, C. *et al.* Preferential expression and function of Toll-like receptor 3 in human astrocytes. *J. Neuroimmunol.* **159,** 12–19 (2005).
27. Taupin, P. & Gage, F. H. Adult neurogenesis and neural stem cells of the central nervous system in mammals. *J. Neurosci. Res.* **69,** 745–749 (2002).
28. Desai, P. & Person, S. Incorporation of the green fluorescent protein into the herpes simplex virus type 1 capsid. *J. Virol.* **72,** 7563–7568 (1998).
29. Reinert, L. S. *et al.* TLR3 deficiency renders astrocytes permissive to herpes simplex virus infection and facilitates establishment of CNS infection in mice. *J. Clin. Invest.* **122,** 1368–1376 (2012).
30. Bieniasz, P. D. Intrinsic immunity: a front-line defense against viral attack. *Nature Immunol.* **5,** 1109–1115 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank our patients, their families and physicians; and the members of the three laboratories for helpful discussions and critical reading of this manuscript. The work was funded by grant number 8UL1TR000043 from the National Center for Translational Sciences (NCATS), the National Institutes of Health (NIH), the Rockefeller University, the St. Giles Foundation, the ANR, INSERM, Paris Descartes University, the March of Dimes, NIH grant 5R01NS072381-02 (to J.-L.C., L.S. and L.D.N.), NIH grant 1R03AI0883502-01 (to L.D.N.), NIH grant 1R01NS066390 and the Manton Foundation, the Israeli Centers of Research Excellence (I-CORE), and Gene Regulation in Complex Human Disease, Center No 41/11 (to I.M.P.). F.G.L. is supported by the New York Stem Cell Foundation.

Author Contributions F.G.L., I.M.P., S.-Y.Z., J.-L.C., L.S. and L.D.N. designed the experiments. F.G.L., I.M.P., S.-Y.Z., M.J.C., M.H., A. A., G.M., S.-W.Y., S.K., P.A.G, J.O.-M., E.J., E.T., Y.E. and T.M.S. carried out the experiments. S.A. and M.T. helped to obtain materials from patients and interpret the findings. G.Q.D. and L.A. helped to analyse and describe the data. S.-Y.Z. and J.-L.C. wrote the manuscript with the aid of F.G.L., I.M.P., L.S. and L.D.N. F.G.L., I.M.P. and S.-Y. Zhang are equal first authors. J.L.C., L.S. and L.D.N. are co-senior authors.

Author Information The transcriptome data have been deposited in the Gene Expression Omnibus database under accession number GSE40593. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.Y.Z. (shzh289@rockefeller.edu) or J.L.C. (jean-laurent.casanova@rockefeller.edu).

LETTER

doi:10.1038/nature11599

Resurrection of endogenous retroviruses in antibody-deficient mice

George R. Young[1], Urszula Eksmond[1], Rosalba Salcedo[2], Lena Alexopoulou[3], Jonathan P. Stoye[4] & George Kassiotis[1]

The mammalian host has developed a long-standing symbiotic relationship with a considerable number of microbial species. These include the microbiota on environmental surfaces, such as the respiratory and gastrointestinal tracts[1], and also endogenous retroviruses (ERVs), comprising a substantial fraction of the mammalian genome[2,3]. The long-term consequences for the host of interactions with these microbial species can range from mutualism to parasitism and are not always completely understood. The potential effect of one microbial symbiont on another is even less clear. Here we study the control of ERVs in the commonly used C57BL/6 (B6) mouse strain, which lacks endogenous murine leukaemia viruses (MLVs) able to replicate in murine cells. We demonstrate the spontaneous emergence of fully infectious ecotropic[4] MLV in B6 mice with a range of distinct immune deficiencies affecting antibody production. These recombinant retroviruses establish infection of immunodeficient mouse colonies, and ultimately result in retrovirus-induced lymphomas. Notably, ERV activation in immunodeficient mice is prevented in husbandry conditions associated with reduced or absent intestinal microbiota. Our results shed light onto a previously unappreciated role for immunity in the control of ERVs and provide a potential mechanistic link between immune activation by microbial triggers and a range of pathologies associated with ERVs, including cancer.

Retroviruses can establish germline infection and become part of the host genome[2,3]. Most, if not all, ERVs have become inactive owing to mutations, or transcriptionally silenced through the action of diverse mechanisms[2,3]. However, RNA and protein expression of replication-defective ERVs are frequently increased in infection, autoimmunity and cancer[2,3]. Whether or not the immune system defends against potential threats posed by ERVs is unclear. To address the role of adaptive immunity in this process, we assessed ERV expression in B6 mice. We initially compared the transcriptional profiles of purified macrophages from B6 wild-type and T- and B-cell-deficient *Rag1*$^{-/-}$ mice. The two transcripts with the highest increase in expression levels in macrophages from *Rag1*$^{-/-}$ mice (Fig. 1a) correspond to the *env* and *gag* genes, respectively (Supplementary Table 1), of an endogenous ecotropic MLV (eMLV) locus, *Emv2*, a replication-defective single-copy ERV present in B6 mice[5]. Differential expression of eMLV was confirmed by quantitative reverse transcriptase PCR (qRT–PCR) for spliced *env* messenger RNA in macrophages (Fig. 1b) and in several tissues (Fig. 1c).

Other ERV families were also differentially expressed in macrophages, albeit less strongly (1.7–2.1-fold; Supplementary Table 1). Not distinguishing between members of multicopy families, expression of polytropic MLVs (pMLVs), xenotropic MLVs (xMLVs) and the *Mus musculus* type D (MusD) retrovirus family of retrotransposons was also increased in the lungs of *Rag1*$^{-/-}$ mice (Supplementary Fig. 1). In line with increased MLV mRNA expression, 60–80% of total splenocytes and all haematopoietic lineages analysed from *Rag1*$^{-/-}$ but not wild-type mice, expressed MLV surface glycoprotein (SU) (Fig. 1d, e).

Tcra$^{-/-}$ or *Tcrd*$^{-/-}$ mice, lacking T-cell receptor (TCR)αβ and TCRγδ T cells, respectively, showed low eMLV expression (Fig. 1f). eMLV expression was similarly low in *H2-A,E*$^{-/-}$ mice (MGI allele

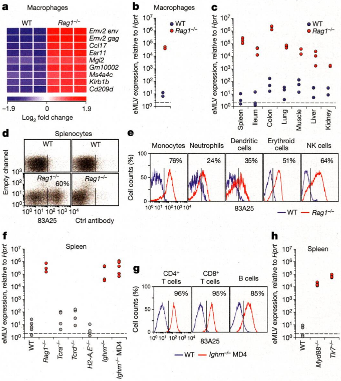


Figure 1 | eMLV activation in antibody-deficient mice. **a**, Significantly upregulated (>4-fold) genes in CD11b^{+} MHC-II$^{\mathrm{hi}}$ B220^{-} Gr1^{-} macrophages from *Rag1*$^{-/-}$ mice compared with macrophages from wild-type (WT) mice. Triplicate microarrays from cells isolated from 40 mice are shown. **b**, eMLV spliced *env* mRNA expression in the same cells as in **a**. Each symbol represents macrophages from 20 mice ($P = 0.024$; paired Student's *t*-test). **c**, eMLV spliced *env* mRNA expression in indicated organs from wild-type or *Rag1*$^{-/-}$ mice (spleen: $P = 0.020$; ileum: $P = 0.032$; colon: $P = 0.004$; lung: $P = 0.001$; muscle: $P = 0.016$; and kidney: $P = 0.009$; unpaired Student's *t*-test). **d**, **e**, MLV SU expression (detected using the 83A25 monoclonal antibody; see Methods) in splenocytes (**d**) or indicated cell types (**e**) from wild-type or *Rag1*$^{-/-}$ mice. **f**, eMLV spliced *env* mRNA expression in the spleens of the indicated strains ($P < 0.001$ between wild-type and either *Ighm*$^{-/-}$ or *Ighm*$^{-/-}$ MD4 mice; one-way analysis of variance (ANOVA)). **g**, MLV SU expression in splenic lymphocytes from wild-type or *Ighm*$^{-/-}$ MD4 mice. **h**, eMLV spliced *env* mRNA expression in the spleens of the indicated strains ($P < 0.001$ between wild-type and either *Myd88*$^{-/-}$ or *Tlr7*$^{-/-}$ mice; one-way ANOVA). In **c**, **f** and **h**, each symbol is an individual mouse. In **d**, **e** and **g**, plots are representative of four mice per group. In **f** and **h**, values above 10^3 were considered high and are indicated by red-filled symbols.

---

[1]Division of Immunoregulation, MRC National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK. [2]Basic Science Program, SAIC-Frederick, Inc. and Cancer and Inflammation Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, Maryland 21701, USA. [3]Centre d'Immunologie de Marseille-Luminy (CIML), Aix-Marseille University-UM2, INSERM-U1104, CNRS-UMR7280, Marseille, France. [4]Division of Virology, MRC National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK.

$H2^{dlAb1\text{-}Ea}$; lacking the region between the *H2-Ab1* and *H2-Ea* genes), deficient in major histocompatibility complex (MHC) class II, MHC II-restricted T cells and T-cell-dependent antibodies (Fig. 1f). By contrast, mice lacking B cells ($Ighm^{-/-}$ mice) or mice unable to produce polyclonal antibodies ($Ighm^{-/-}$ MD4 mice) expressed substantially higher eMLV levels than did wild-type mice (Fig. 1f), demonstrating that high eMLV expression characterized mice lacking antigen-specific antibodies. In addition to splenocytes from $Rag1^{-/-}$ mice (Fig. 1e), T and B cells from $Ighm^{-/-}$ MD4 mice, but not from wild-type control mice, expressed MLV SU (Fig. 1g), indicating that eMLV can also be highly expressed in lymphocytes.

Toll-like receptors (TLRs) have been implicated in the control of B-cell responses, and $Myd88^{-/-}$ and $Tlr7^{-/-}$ mice have significantly reduced serum levels of natural antibodies and a defective antibody response to immunization or infection, including with retroviruses[6–9]. Notably, the expression of eMLV was markedly increased in $Myd88^{-/-}$ and $Tlr7^{-/-}$ mice in comparison with wild-type mice (Fig. 1h). Similarly increased eMLV expression was observed in $Tlr7^{-/-}$ mice, but not in $Tlr9^{-/-}$ mice housed in a different facility (Supplementary Fig. 2).

To investigate the mechanistic link between antibody deficiencies and increased MLV expression, we examined the origin of eMLV transcription. The B6 genome does not contain replication-competent eMLV proviruses, and, although the *Emv2* locus can produce mRNA, it is unable to produce infectious virus owing to an inactivating G-to-C mutation at position 3576 of the *pol* region[5,10]. In addition, *Emv2 gag* encodes an N-tropic capsid, which would be restricted by the $Fv1^{b}$ restriction factor in B6 mice[10]. However, it was theoretically possible that recombination between replication-defective *Emv2* and non-ecotropic MLVs resulted in an MLV with full infectivity[11] that could spread in $Rag1^{-/-}$ mice. Remarkably, the plasmas of young and old $Rag1^{-/-}$ mice, but not of wild-type control mice, contained retroviruses that were capable of replicating in mouse cells *in vitro* (Fig. 2a), which we refer to as $Rag1^{-/-}$ mouse-associated retroviruses (RARVs). Sequencing of the *pol* region demonstrated repair of the *Emv2*-inactivating mutation in all RARV isolates (Supplementary Fig. 3). Functional *in vitro* assays (Fig. 2b) and sequencing of the *gag* region (Fig. 2c) showed that RARVs also exhibited B-tropism. Genome sequence comparisons between RARVs showed that young $Rag1^{-/-}$ mice contained highly similar viruses, which diverged substantially in old $Rag1^{-/-}$ mice (Fig. 2d). All RARVs were recombinants between *Emv2* and endogenous non-ecotropic MLVs (Supplementary Fig. 4). The *pol* defect of *Emv2* was probably restored in RARVs by recombination with *Xmv43* (also known as *Bxv1*; Supplementary Fig. 4), an ERV that contains a functional *pol* region but is unable to infect mouse cells owing to polymorphisms in the mouse cellular receptor[2]. Recombination events involving *Xmv43* have also been found to be responsible for the emergence of leukaemogenic MLVs in AKR mice[12]. However, the switch in capsid tropism resulted from recombination with other endogenous xMLVs (Supplementary Fig. 4). Notably, the divergence of RARVs isolated from old $Rag1^{-/-}$ mice was due to further recombination replacing the ecotropic *env* with polytropic *env* from *Pmv1*, *Pmv5* or *Pmv16* (Supplementary Fig. 4). Together, these findings indicate the emergence of infectious eMLVs that could have infected $Rag1^{-/-}$ mice and given further rise to infectious pMLVs. Supporting this notion, most *gag/pol* eMLV mRNA detected in $Rag1^{-/-}$ mice seemed to be transcribed from integrated RARVs, rather than the germline copy of *Emv2* (Supplementary Fig. 5).

Emv2 and non-ecotropic MLV recombination events resulting in infectious eMLV generation might occur *de novo* in individual $Rag1^{-/-}$ mice. Alternatively, these RARVs might have been vertically transmitted through successive generations. To test the latter possibility directly and assess the capacity of RARVs to spread in immunodeficient mice, we established a colony of $Rag1^{-/-}$ $Emv2^{-/-}$ mice by first crossing an $Emv2^{-/-}$ male mouse[13] with a $Rag1^{-/-}$ female mouse, and then intercrossing selected progeny to homozygosity. These mice lack the germline copy of *Emv2*, meaning any infectious eMLV present would have been vertically transmitted. Both eMLV spliced *env* mRNA (Fig. 2e) and MLV SU expression (Fig. 2f) were readily detected in the spleens of $Rag1^{-/-}$ $Emv2^{-/-}$ mice in this colony. Furthermore, analysis of eMLV *env* and *pol* DNA copies indicated extensive replication of vertically transmitted RARVs in $Rag1^{-/-}$ $Emv2^{-/-}$ mice (Fig. 2g).

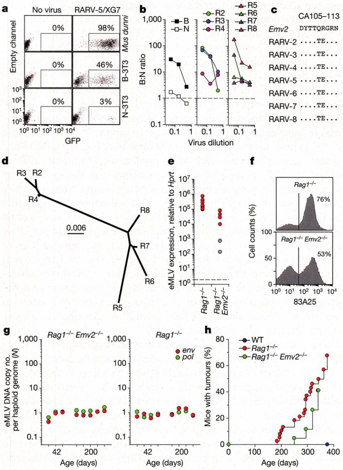


Figure 2 | Retroviraemia and leukaemias/lymphomas in antibody-deficient mice. a, Detection of infectious MLV (RARV-5/XG7) from the plasma of a representative $Rag1^{-/-}$ mouse by restoring infectivity of the green fluorescent protein (GFP)-expressing XG7 retroviral vector in the indicated cell type. Numbers within the plots denote the percentage of retrovirally transduced (GFP^{+}) cells. **b**, Fv1 tropism of RARVs isolated from 6 (R2–R4)- or 25 (R5–R8)-week-old healthy $Rag1^{-/-}$ mice, shown as the ratio of infectivity in B-3T3 to N-3T3 cells (B:N ratio). B- and N-tropic strains of Friend MLV (F-MLV) are shown for comparison. **c**, Amino acid residues of capsid positions 105–113 (CA105–113) deduced from the nucleotide sequence of *Emv2* and the same RARVs as in **b**. Dots indicate identities. **d**, Phylogenetic tree of the same RARVs as in **b**. The scale indicates the probability of base substitution per site. **e**, eMLV spliced *env* mRNA expression in the spleens of $Rag1^{-/-}$ mice or vertically infected $Rag1^{-/-}$ $Emv2^{-/-}$ mice. Each dot is an individual mouse ($P < 0.001$; unpaired Student's *t*-test). Values above 10^{3} were considered high and are indicated by red-filled symbols. **f**, MLV SU expression in splenocytes from $Rag1^{-/-}$ or vertically infected $Rag1^{-/-}$ $Emv2^{-/-}$ mice (representative of nine mice per group). **g**, eMLV DNA copy numbers per haploid genome, determined by qPCR for the *pol* or ecotropic *env* gene, in DNA from the spleens of healthy $Rag1^{-/-}$ mice (right) or vertically infected $Rag1^{-/-}$ $Emv2^{-/-}$ mice (left). Symbols represent individual mice, grouped according to their age. The sensitivity limit of this PCR method was determined as a median of 0.0003 copies per haploid genome, using $Emv2^{-/-}$ mice. eMLV DNA copy numbers for $Rag1^{-/-}$ mice include *Emv2* (1/*N*). **h**, Tumour (leukaemias/lymphomas) incidence in cohorts of wild-type ($n = 37$), $Rag1^{-/-}$ ($n = 38$) or vertically infected $Rag1^{-/-}$ $Emv2^{-/-}$ mice ($n = 23$) at the NIMR SPF facility ($P < 0.000001$ between wild-type and $Rag1^{-/-}$ mice; $P = 0.00025$ between wild-type and $Rag1^{-/-}$ $Emv2^{-/-}$ mice; log-rank survival analysis).

By contrast, sexual or *in utero* infection was not observed in separate crosses of either male or female virus-positive *Rag1*$^{-/-}$ *Emv2*$^{-/-}$ mice with virus-free *Emv2*$^{-/-}$ mice (Supplementary Fig. 6).

To examine the potential of RARVs to replicate in *Rag1*$^{-/-}$ mice further, we assessed the frequency of tumours characteristic of retroviral infection[2,3] in cohorts of *Rag1*$^{-/-}$ mice. Notably, starting from 180 days and affecting 67% of the animals by 380 days, *Rag1*$^{-/-}$ mice, but not wild-type control mice, showed signs of morbidity (Fig. 2h). On examination, large tumours, often associated with anaemia, were observed in all morbid *Rag1*$^{-/-}$ mice (Fig. 2h and Supplementary Fig. 7). The pathogenic potential of infection with RARVs was established in aged cohorts of vertically infected *Rag1*$^{-/-}$ *Emv2*$^{-/-}$ mice, which developed tumours at a comparable incidence rate (Fig. 2h).

Thymic or splenic tumours in *Rag1*$^{-/-}$ mice consisted mainly of a single MLV SU-expressing cell type, which differed between animals, and had the histological appearance of lymphoblastic lymphosarcomas (Supplementary Fig. 8). Discrete chromosomal aberrations were found in most tumours analysed (Supplementary Fig. 8), suggestive of clonal origin. Consistent with MLV production by tumour cells, we observed an abundance of MLV-type particles in the extracellular space of tumour samples, but not in a spleen sample from a healthy *Rag1*$^{-/-}$ mouse (Supplementary Fig. 9). Furthermore, a substantial increase in both eMLV *env* and *pol* DNA copy numbers was detected in all tumour samples, with one exception in which only *pol* DNA copies were increased (Supplementary Fig. 9), indicating that RARVs had extensively infected the cells that gave rise to lymphomas. Together, these results support a model in which several recombination events restore *Emv2* infectivity, leading to spontaneous retroviraemia and vertical transmission to progeny, and eventually drive an oncogenic process similar to that extensively described in mouse strains carrying fully infectious ERVs[2,3].

Our results associated a lack of antibodies with establishment of infectious eMLVs in mouse colonies. Next, we investigated the potential mode of antibody action. Antiretroviral antibodies have a long-established role in limiting the spread of infectious endogenous retroviruses[14], both within and between animals. However, it was also possible that antibodies were preventing a step before the emergence of infectious eMLV recombinants. Rescue of *Emv2* infectivity by recombination with a non-ecotropic MLV necessitates co-expression of both proviruses in the same cell at sufficient levels for co-packaging into the same virion. Low expression of these proviruses in wild-type mice could be a rate-limiting factor in the emergence of infectious eMLVs. However, expression of certain endogenous MLVs in mouse cells is known to be inducible, notably, by microbial products[2,3,15]. For example, bacterial lipopolysaccharide (LPS) stimulation activates non-ecotropic MLVs, and *Xmv43* in particular[15–19]. To examine the responsiveness of ERVs and other retroelements to microbial stimulation, we took advantage of probes or probe sets in standard microarray platforms that report ERV or retroelement expression. Analysis of a publicly available data set[20] uncovered specific induction of non-ecotropic MLV transcripts by LPS and polyinosinic:polycytidylic acid (poly(I:C)), and suppression by Pam$_3$CSK$_4$ (Fig. 3a, b). Early transposon transcripts were also induced by poly(I:C) (Fig. 3a). These findings were confirmed by LPS stimulation, which induced high MLV SU expression in wild-type and *Emv2*$^{-/-}$ mouse splenocytes, also in a MyD88-independent manner (Fig. 3c and Supplementary Fig. 10). These *in vitro* conditions did not significantly induce eMLV expression (Fig. 3a). However, *ex vivo* analysis showed higher expression of eMLV (Fig. 1c), as well as of other ERVs/retroelements (Supplementary Fig. 1), specifically in the colon, suggestive of microbial involvement.

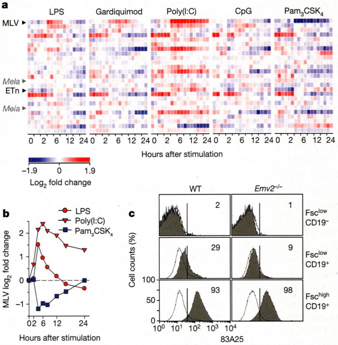


Figure 3 | Mouse ERV activation by microbial products. **a**, ERV/retroelement-reporting probe set (Supplementary Table 3) signals in a publicly available Affymetrix HT mouse genome 430A microarray data set[20] (ArrayExpress accession E-GEOD-17721) of wild-type B6 bone marrow-derived dendritic cells after stimulation with microbial products. Black arrows indicate the probe sets that are significantly regulated ($P < 0.05$) more than twofold by at least one stimulus. *Emv2*-specific probe sets (annotated as *Mela*) are also indicated by grey arrows for comparison. ETn, early transposon. **b**, Mean log$_2$ fold change in the MLV-reporting probe set in the same data set. **c**, MLV SU expression in wild-type or *Emv2*$^{-/-}$ splenocytes before (open histograms) or after (filled histograms) stimulation with 10 μg ml^{-1} LPS for 48 h, according to forward scatter (Fsc) and CD19 expression. Numbers in the plots denote the percentage of cells within each gate and represent two donors each analysed in duplicate.

Antibodies have established roles in controlling intestinal bacteria and neutralizing their products, such as LPS, in the gut lumen or the systemic circulation, and antibody-deficient mice are known to display increased microbial trasnslocation[1,9,21–23]. It was, therefore, possible that antibody deficiency allowed microbial products to induce expression of MLV proviruses in *Rag1*$^{-/-}$ mice, including the parents of recombinant RARVs (Supplementary Fig. 1). In support of this notion, and in agreement with the established role of natural IgM in systemic clearance of bacterial LPS and protection from endotoxaemia[9,21,22], production of non-hypermutated IgM alone was sufficient for eMLV control (Supplementary Fig. 11). If antibodies required for preventing MLV expression were, indeed, against several microbial products, then antibody deficiency should not result in increased MLV expression in the absence of microbial triggers (Supplementary Fig. 12).

To begin to examine the contribution of microbial triggers, we measured eMLV expression in specific pathogen-free (SPF) *Rag1*$^{-/-}$ mice from colonies that differed in intestinal microbiota. Importantly, the use of embryo transfer for the rederivation of these independent colonies removes adventitious organisms, including vertically transmitted eMLVs. Therefore, any RARVs found in these rederived *Rag1*$^{-/-}$ mouse colonies must be generated *de novo* in the life-history of each colony. In stark contrast to *Rag1*$^{-/-}$ mice that were maintained on neutral pH water at the National Institute for Medical Research (NIMR), colonies of *Rag1*$^{-/-}$ mice that were maintained on acidified water expressed minimal eMLV levels (Fig. 4a). Water acidification reduced overall bacterial diversity in the colons of *Rag1*$^{-/-}$ mice (Supplementary Fig. 13 and Supplementary Table 2) and is a common precautionary measure used in many animal facilities that reduces bacterial colonization within the intestinal tract and translocation into the circulation[24]. Lack of eMLV expression was noted in *Rag1*$^{-/-}$ mice obtained from the Jackson Laboratory (JAX), also maintained on acidified water (Fig. 4a). Furthermore, minimal levels of eMLV were detected in *Rag1*$^{-/-}$ mice on neutral pH water at the Rodent Center

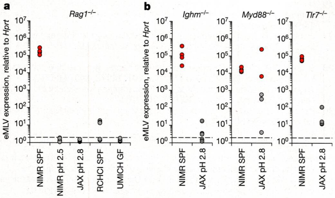


Figure 4 | eMLV activation in antibody-deficiency depends on husbandry conditions. **a**, eMLV spliced *env* mRNA expression in the spleens of *Rag1*$^{-/-}$ mice on neutral pH (SPF) or acidified water (pH 2.5) at the NIMR, on acidified water (pH 2.8) at JAX, on neutral pH at RCHCI, or in germ-free (GF) facilities at UMICH ($P < 0.016$ between *Rag1*$^{-/-}$ mice at SPF NIMR and all other groups; one-way ANOVA). **b**, eMLV spliced *env* mRNA expression in the spleens of *Ighm*$^{-/-}$, *Myd88*$^{-/-}$ or *Tlr7*$^{-/-}$ mice on neutral pH water (SPF) at the NIMR or on acidified water (pH 2.8) at JAX ($P = 0.005$ and $P = 0.029$ for *Ighm*$^{-/-}$ and *Tlr7*$^{-/-}$ mice, respectively; unpaired Student's *t*-test). Each dot is an individual mouse and values above 10^3 were considered high and are indicated by red-filled symbols.

HCI (RCHCI; Fig. 4a), which contained distinct bacterial genera in comparison with *Rag1*$^{-/-}$ mice at NIMR (Supplementary Fig. 13 and Supplementary Table 2). Lastly, negligible levels of eMLV were expressed in *Rag1*$^{-/-}$ mice in germ-free facilities at the University of Michigan (UMICH) and offered neutral pH water (Fig. 4a). The latter two conditions also distinguished between effects of acidified water on intestinal flora and effects on other physiological processes. Thus, high eMLV expression in independently rederived colonies of *Rag1*$^{-/-}$ mice correlated with the presence of the normal SPF microbiota.

Husbandry conditions contributed to high eMLV expression also in independently rederived strains with distinct immunodeficiencies. Colonies of *Ighm*$^{-/-}$ and *Tlr7*$^{-/-}$ mice maintained on acidified water at JAX expressed minimal levels of eMLV (Fig. 4b). Variable eMLV expression levels were detected in *Myd88*$^{-/-}$ mice at JAX (Fig. 4b), probably because, once generated, vertical transmission of infectious eMLVs was unaffected by water acidification. Although defining the precise role of the microbiota on eMLV induction will require further investigation, collectively our results demonstrate that the high eMLV expression phenotype, in genotypes causing immunodeficiency, requires environmental interaction.

High eMLV expression in independently rederived immunodeficient colonies reveals *de novo* eMLV induction in each colony. It does not, however, indicate when or how often in the life-history of a colony infectious eMLVs may emerge. Adult RARV-free *Rag1*$^{-/-}$ mice, previously on acidified water, maintained low eMLV expression after a switch to neutral pH water, indicating a low probability of RARV emergence and spread in an individual mouse. To examine whether this probability is low in general or could be higher during early mouse development, we monitored successive generations of *Ighm*$^{-/-}$ mouse colonies from two independent rederivations, one of which was recent, into the NIMR SPF facility. This analysis suggested that eMLV induction occurred during the first few filial generations (F; Supplementary Fig. 14). Moreover, eMLV-expressing *Myd88*$^{-/-}$ mice at JAX (Fig. 4b), were at F_5 of homozygous breeding. Thus, although low in individual mice, there is high cumulative probability of infectious eMLV emergence, involving a sequence of recombination events similar to those seen in *Rag1*$^{-/-}$ mice (Supplementary Fig. 12), and subsequent establishment in an antibody-deficient colony over a few generations.

The B6 strain has historically dominated many research fields, partly owing to the resistance of this strain to retrovirally induced tumours. Our results demonstrate that this important attribute of B6 mice is conditional on their immune competence, with significant implications both for the design and interpretation of mouse studies. Furthermore, our results show that ERV activation is determined by husbandry conditions, thus accentuating potential differences in ERV expression between animal facilities.

Although well-established in the mouse[2,3], the oncogenic potential of ERVs in humans has not been observed[25,26]. However, non-long terminal repeat (LTR) retroelement families have been documented to have caused human cancers by insertional mutagenesis at the somatic level[25,26]. Interestingly, we found that TLR stimulation of human cells induced expression of distinct ERVs and retroelements, including the mammalian apparent LTR retrotransposon (MaLR) family (Supplementary Fig. 15), previously implicated in the pathogenesis of human lymphomas[27]. Transcription of human ERVs and retroelements can also be induced by physiological activation of both adaptive and innate immune cells[28]. Moreover, increased risk of lymphomas in humans is linked to infection or inflammation[29] and also to antibody deficiencies[30]. Thus, interactions between microbial symbionts, leading to ERV/retroelement activation, may provide a mechanistic link between cancer and stimulation of the immune system by microbiota or pathogenic infections.

METHODS SUMMARY

Mice. Inbred C57BL/6J (B6) or genetically modified B6-backcrossed mice were from the following facilities: NIMR (London, UK); JAX (Massachusetts, USA); UMICH (Michigan, USA); CIML (Marseille, France); London Research Institute, Cancer Research UK (London, UK); RCHCI (Zürich, Switzerland); or Center for Cancer Research, National Cancer Institute (Maryland, USA). Mice were kept in individually ventilated cages in SPF facilities or in germ-free isolators. Mice were maintained on either neutral pH or acidified water. Where indicated, colonies were established by rederivation into each of these facilities. This process ensures the removal of adventitious agents and leads to colonization of the newly rederived mice with the specific microbiota of each facility (or the lack of colonization in the case of the germ-free facility). All animal experiments were approved by ethical committees of respective institutes, and conducted according to local guidelines and regulations.

Full Methods and any associated references are available in the online version of the paper.

Received 8 June; accepted 18 September 2012.

Published online 24 October; corrected online 28 November 2012 (see full-text HTML version for details).

1. Honda, K. & Littman, D. R. The microbiome in infectious disease and inflammation. *Annu. Rev. Immunol.* **30,** 759–795 (2012).
2. Stocking, C. & Kozak, C. A. Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65,** 3383–3398 (2008).
3. Stoye, J. P. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nature Rev. Microbiol.* **10,** 395–406 (2012).
4. Stoye, J. P. & Coffin, J. M. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. *J. Virol.* **61,** 2659–2669 (1987).
5. King, S. R., Berson, B. J. & Risser, R. Mechanism of interaction between endogenous ecotropic murine leukemia viruses in (BALB/c × C57BL/6) hybrid cells. *Virology* **162,** 1–11 (1988).
6. Demaria, O. *et al.* TLR8 deficiency leads to autoimmunity in mice. *J. Clin. Invest.* **120,** 3651–3662 (2010).
7. DeFranco, A. L., Rookhuizen, D. C. & Hou, B. Contribution of Toll-like receptor signaling to germinal center antibody responses. *Immunol. Rev.* **247,** 64–72 (2012).
8. Browne, E. P. Regulation of B cell responses by Toll-like receptors. *Immunology* **136,** 370–379 (2012).
9. Kirkland, D. *et al.* B cell-intrinsic MyD88 signaling prevents the lethal dissemination of commensal bacteria during colonic damage. *Immunity* **36,** 228–238 (2012).
10. Li, M., Huang, X., Zhu, Z. & Gorelik, E. Sequence and insertion sites of murine melanoma-associated retrovirus. *J. Virol.* **73,** 9178–9186 (1999).
11. Pothlichet, J., Mangeney, M. & Heidmann, T. Mobility and integration sites of a murine C57BL/6 melanoma endogenous retrovirus involved in tumor progression *in vivo*. *Int. J. Cancer* **119,** 1869–1877 (2006).
12. Stoye, J. P., Moroni, C. & Coffin, J. M. Virological events leading to spontaneous AKR thymomas. *J. Virol.* **65,** 1273–1285 (1991).

13. Young, G. R. *et al.* Negative selection by an endogenous retrovirus promotes a higher-avidity CD4$^+$ T cell response to retroviral infection. *PLoS Pathog.* **8,** e1002709 (2012).
14. Melamedoff, M., Lilly, F. & Duran-Reynals, M. L. Suppression of endogenous murine leukemia virus by maternal resistance factor. *J. Exp. Med.* **158,** 506–514 (1983).
15. Stoye, J. P. & Moroni, C. Endogenous retrovirus expression in stimulated murine lymphocytes. *J. Exp. Med.* **157,** 1660–1674 (1983).
16. Kozak, C. A. & Rowe, W. P. Genetic mapping of xenotropic murine leukemia virus-inducing loci in five mouse strains. *J. Exp. Med.* **152,** 219–228 (1980).
17. McCubrey, J. & Risser, R. Genetic interactions in induction of endogenous murine leukemia virus from low leukemic mice. *Cell* **28,** 881–888 (1982).
18. Moroni, C. & Schumann, G. Lipopolysaccharide induces C-type virus in short term cultures of BALB/c spleen cells. *Nature* **254,** 60–61 (1975).
19. Greenberger, J. S., Phillips, S. M., Stephenson, J. R. & Aaronson, S. A. Induction of mouse type-C RNA virus by lipopolysaccharide. *J. Immunol.* **115,** 317–320 (1975).
20. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326,** 257–263 (2009).
21. Lim, A. *et al.* Antibody and B-cell responses may control circulating lipopolysaccharide in patients with HIV infection. *AIDS* **25,** 1379–1383 (2011).
22. Reid, R. R. *et al.* Endotoxin shock in antibody-deficient mice: unraveling the role of natural antibody and complement in the clearance of lipopolysaccharide. *J. Immunol.* **159,** 970–975 (1997).
23. Shulzhenko, N. *et al.* Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nature Med.* **17,** 1585–1593 (2011).
24. Wu, L. *et al.* Chronic acid water feeding protects mice against lethal gut-derived sepsis due to *Pseudomonas aeruginosa*. *Curr. Issues Intest. Microbiol.* **7,** 19–28 (2006).
25. Belancio, V. P., Roy-Engel, A. M. & Deininger, P. L. All y'all need to know 'bout retroelements in cancer. *Semin. Cancer Biol.* **20,** 200–210 (2010).
26. Romanish, M. T., Cohen, C. J. & Mager, D. L. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. *Semin. Cancer Biol.* **20,** 246–253 (2010).
27. Lamprecht, B. *et al.* Derepression of an endogenous long terminal repeat activates the *CSF1R* proto-oncogene in human lymphoma. *Nature Med.* **16,** 571–579 (2010).
28. Bannert, N. & Kurth, R. Retroelements and the human genome: new perspectives on an old relation. *Proc. Natl Acad. Sci. USA* **101,** 14572–14579 (2004).
29. Trinchieri, G. Cancer and inflammation: an old intuition with rapidly evolving new concepts. *Annu. Rev. Immunol.* **30,** 677–706 (2012).
30. Park, M. A. *et al.* Common variable immunodeficiency: a new look at an old disease. *Lancet* **372,** 489–502 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We wish to thank W.-D. Hardt for mouse samples and discussion, L. Sellés Vidal for technical assistance, and colleagues for critical reading of the manuscript. We also wish to thank the staff of the Unit for Laboratory Animal Medicine, University of Michigan, for the provision of germ-free mice. We are grateful for assistance from the Division of Biological Services, the Flow Cytometry, Electron Microscopy and Microarray facilities at NIMR. This work was supported by the UK Medical Research Council (U117581330 and U117512710).

Author Contributions G.R.Y., J.P.S. and G.K. designed the study. G.R.Y. and U.E. carried out experiments and analysed data. R.S. and L.A. provided data or study samples. G.R.Y., J.P.S. and G.K. prepared the manuscript.

Author Information Primary microarray data from triplicate arrays were deposited at ArrayExpress under accession E-MEXP-3623. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.K. (gkassio@nimr.mrc.ac.uk).

LETTER

doi:10.1038/nature11580

Structure of the chemokine receptor CXCR1 in phospholipid bilayers

Sang Ho Park[1], Bibhuti B. Das[1], Fabio Casagrande[1], Ye Tian[1,2], Henry J. Nothnagel[1], Mignon Chu[1], Hans Kiefer[3], Klaus Maier[4], Anna A. De Angelis[4], Francesca M. Marassi[2] & Stanley J. Opella[1]

CXCR1 is one of two high-affinity receptors for the CXC chemokine interleukin-8 (IL-8), a major mediator of immune and inflammatory responses implicated in many disorders, including tumour growth[1–3]. IL-8, released in response to inflammatory stimuli, binds to the extracellular side of CXCR1. The ligand-activated intracellular signalling pathways result in neutrophil migration to the site of inflammation[2]. CXCR1 is a class A, rhodopsin-like G-protein-coupled receptor (GPCR), the largest class of integral membrane proteins responsible for cellular signal transduction and targeted as drug receptors[4–7]. Despite its importance, the molecular mechanism of CXCR1 signal transduction is poorly understood owing to the limited structural information available. Recent structural determination of GPCRs has advanced by modifying the receptors with stabilizing mutations, insertion of the protein T4 lysozyme and truncations of their amino acid sequences[8], as well as addition of stabilizing antibodies and small molecules[9] that facilitate crystallization in cubic phase monoolein mixtures[10]. The intracellular loops of GPCRs are crucial for G-protein interactions[11], and activation of CXCR1 involves both amino-terminal residues and extracellular loops[2,12,13]. Our previous nuclear magnetic resonance studies indicate that IL-8 binding to the N-terminal residues is mediated by the membrane, underscoring the importance of the phospholipid bilayer for physiological activity[14]. Here we report the three-dimensional structure of human CXCR1 determined by NMR spectroscopy. The receptor is in liquid crystalline phospholipid bilayers, without modification of its amino acid sequence and under physiological conditions. Features important for intracellular G-protein activation and signal transduction are revealed. The structure of human CXCR1 in a lipid bilayer should help to facilitate the discovery of new compounds that interact with GPCRs and combat diseases such as breast cancer.

To examine the structure and function of CXCR1 in its natural environment, we reconstituted the full-length, active receptor in phospholipid bilayers (proteoliposomes). The NMR method we developed, rotationally aligned (RA) solid-state NMR[15], is specifically tailored for the unique properties of membrane proteins in liquid crystalline phospholipid bilayers. It combines features of magic angle spinning (MAS)[16] and oriented-sample (OS)[17] solid-state NMR to resolve and assign resonances associated with each amino acid residue, measure site-specific orientation restraints relative to the bilayer, and calculate the three-dimensional structure of the protein and its integral membrane orientation. RA solid-state NMR differs from previously used OS methods because it relies on the inherent rotational diffusion of membrane proteins in phospholipid bilayers[18] to provide orientation-dependent motional averaging of dipolar coupling (DC) powder patterns relative to the bilayer normal, rather than the orientation-dependent frequencies of single-line resonances observed in OS NMR of stationary, uniaxially aligned samples. Furthermore, the method takes advantage of recent bioinformatics developments that facilitate molecular fragment replacement approaches to structure determination, including membrane proteins[19–21].

CXCR1 was uniformly $^{13}C/^{15}N$-labelled by expression in *Escherichia coli*, then purified, refolded in 1,2-dimyristoyl-*sn*-glycero-3-phosphatidylcholine (DMPC) proteoliposomes[22,23], and placed as a concentrated suspension in an MAS rotor. Refolded CXCR1 binds IL-8 with high affinity (dissociation constant (K_d) $\approx$ 1–5 nM) and couples to its G protein, $G_{i/o}$ (half-maximum effective concentration (EC_{50}) = 1 nM)[22], indicating that the NMR sample conditions are compatible with physiological activity (Supplementary Fig. 1).

As expected, the two-dimensional $^{13}C/^{13}C$ correlation spectrum of this 350-residue protein is quite crowded (Supplementary Fig. 2). However, expansion of the $^{13}C\alpha$ resonance region has sufficient resolution to contribute to the assignment process (Supplementary Fig. 2d). Spectra, obtained from a uniformly ^{13}C-labelled sample with a shorter mixing time and from a sample labelled using 2-^{13}C-glycerol, have fewer signals and improved resolution. Regardless, the vast majority of data used to resolve, assign and measure isotropic chemical shift frequencies from N, Cα, CO and Cβ sites were obtained from ^{13}C-detected three-dimensional, triple-resonance experiments (Fig. 1b, Supplementary Figs 3 and 4 and Supplementary Table 2). Overall, 97% of the backbone resonances for residues 20 to 325 were assigned. The missing resonances are from seven Pro residues (Pro 22, Pro 93, Pro 170, Pro 180, Pro 185, Pro 214 and Pro 257) and one Arg residue (Arg 285). None of the ^{15}N and ^{13}C signals from the mobile amino and carboxy termini (residues 1–19 and 326–350) could be detected in the spectra, consistent with our observation of these signals in solid-state NMR experiments designed to detect only signals from mobile sites (Supplementary Fig. 6), as well as our previous analysis of local and global motions of CXCR1 (ref. 24).

Three-dimensional ^{13}C-detected separated local field (SLF) experiments[15] were used to measure the ^1H-^{15}N DC and ^1H-$^{13}C\alpha$ DC frequencies that provide orientation restraints for structure determination (Fig. 1c, d and Supplementary Fig. 5). The protein backbone structure was calculated by a molecular fragment replacement approach. An initial structural model was generated from a set of molecular fragments generated with CS-Rosetta[20] from the experimental chemical shifts, the amino acid sequence of CXCR1, and the helical framework of the prototypical class A GPCR bovine rhodopsin[25]. This initial model was first refined with the experimental restraints using the all atom[19] and the implicit membrane[21] potentials of Rosetta. Finally, the resulting structural model was refined by restrained simulated annealing using Xplor-NIH[26].

The three-dimensional structure of CXCR1 (Fig. 1e) has the consensus fold of a GPCR, with seven transmembrane helices (TM1–TM7) connected by three extracellular loops (ECL1–ECL3) and three intracellular loops (ICL1–ICL3). The average backbone pairwise root mean squared deviation (r.m.s.d.) is 1.7 Å (Supplementary Table 1), and the experimentally measured ^1H-^{15}N DC and ^1H-$^{13}C\alpha$ DC values correlate remarkably well with those calculated from the refined protein structure

---

[1]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0307, USA. [2]Sanford Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. [3]HBC Hochschule Biberach, Karlstrasse 11, 88400 Biberach, Germany. [4]Membrane Receptor Technologies, San Diego, California 92121-3832, USA.

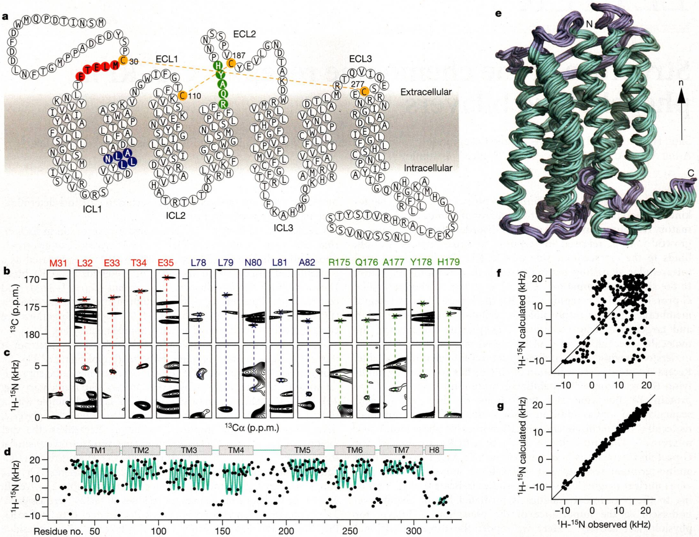


Figure 1 | Structure determination of CXCR1. **a**, CXCR1 topology with two disulphide bonds (gold). **b**, **c**, Strip plots from three-dimensional experiments taken at specific ^{15}N- and ^{13}Cα-chemical shifts for three representative regions of CXCR1: N terminus (residues 31–35) (red), TM2 (residues 78–82) (blue), and ECL2 (residues 175–179) (green). **b**, NCACX data used for resonance assignments. **c**, ^{13}C-detected ^{1}H-^{15}N SLF spectra. **d**, Dipolar wave plot of the experimentally measured ^{1}H-^{15}N DC values as a function of residue number. Sinusoidal fits (cyan) to the data (4.1 kHz r.m.s.d.) highlight the transmembrane (TM1–TM7) and C-terminal (H8) helices. **e**, Ensemble of the 10 lowest energy structures of CXCR1 aligned in the membrane (n denotes bilayer normal). **f**, **g**, Correlation plots of experimental and back-calculated ^{1}H-^{15}N DC restraints obtained before (**f**) and after (**g**) refinement against the experimental data.

(Fig. 1g and Supplementary Fig. 7). Notably, the correlations improve markedly after refinement of the initial structural model with the experimental data, demonstrating that the NMR structure of CXCR1 is determined by the experimentally measured backbone orientation restraints and dihedral angles. We anticipate that both structural accuracy and precision will improve with inclusion of side chain restraints and, where feasible, distance restraints.

DC values contain information about molecular orientation as well as dynamics, and scaling of their values by local internal motions would compromise their analysis in terms of pure orientation restraints. For CXCR1, analysis of the DC data yielded similar values of the magnitude and symmetry of the molecular order tensor for residues in the helices and loops, indicating that these regions of the protein experience a similar degree of order in the lipid bilayer. This is supported by the excellent fit of the DC correlation plots obtained with a single value of the order tensor (Fig. 1g and Supplementary Fig. 7b) and by the observation that more than 20% of the ^{1}H-^{15}N DC signals from residues distributed throughout the protein sequence, including in loop sites, fall within 10% of the maximum value (21 kHz) expected for a static crystalline sample (Fig. 1d).

Following the structure determination of rhodopsin from three- and two-dimensional crystals[25,27], the structures of several class A ligand-activated GPCRs have recently been determined by X-ray crystallography[6,7]. CXCR1 is now the first, to our knowledge, GPCR with its structure determined in liquid crystalline phospholipid bilayers, and the first ligand-activated GPCR with its structure determined without modification of its amino acid sequence (Fig. 2). The structure of CXCR1 shares significant similarities with that of CXCR4 (ref. 28), the only other chemokine receptor of which the structure has been determined (Supplementary Fig. 8). However, there are some notable differences reflecting the modifications made to the sequence of CXCR4 required for crystallization (insertion of T4 lysozyme in ICL3, removal of 33 C-terminal residues and Leu125Trp mutation), the amino acid sequence differences between the two proteins, the influence of the planar phospholipid bilayer, or the rotational diffusion of the protein.

The CXCR1 helices are well defined by the spectroscopic data. For example, a plot of the measured values of the amide ^{1}H-^{15}N DC versus residue number yields a characteristic wave-like pattern[29] reflecting helical structure (Fig. 1d), with breaks in the waves corresponding to

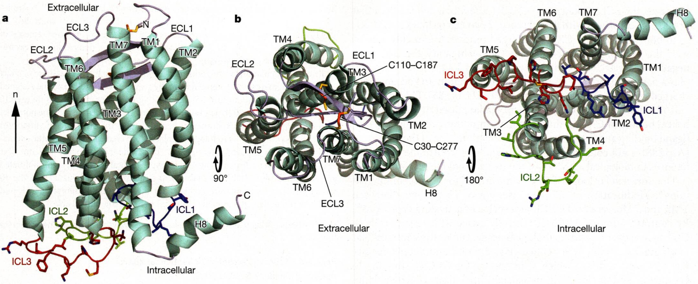


Figure 2 | Three-dimensional structure of CXCR1. Backbone representation of CXCR1 showing helices in cyan (TM1–TM7 and H8), extracellular loops in grey (ECL1–ECL3), and intracellular loops in blue (ICL1), green (ICL2) and red (ICL3). Disulphide-bonded Cys pairs (Cys 30–Cys 277 and Cys 110–Cys 187) are shown as sticks. **a**, Side view (n denotes bilayer normal). **b**, View from the extracellular side. **c**, View from the intracellular side.

helix termini or kinks. For example, the DC data show the presence of a kink that changes the direction of helix TM2 (residues 74–101) at Phe 88, coinciding with a kink at the same location in CXCR4 (Fig. 3). The extracellular start of TM7, just after ECL3, is tilted towards the central axis of the receptor in CXCR1, although it is less well defined in CXCR1 than in CXCR4, in which residues in the N terminus of TM7 interact with the added compound IT1t (ref. 28). Helix TM7 is also about one turn longer at the intracellular end than its counterpart in CXCR4, extending to Ile 308, three residues beyond the conserved GPCR sequence Asn-Pro-X-X-Tyr (in which X denotes any amino acid). Furthermore, residues immediately preceding the mobile C terminus of CXCR1 form a well-defined helix (H8; residues Gln 310 to Ala 321) that is absent in the structure of CXCR4 (Figs 2 and 3 and Supplementary Fig. 8). H8 has a distinctly amphipathic amino acid sequence and aligns along the membrane surface, indicating that the phospholipid bilayer may have a role in stabilizing its conformation.

The NMR data further show the presence of two disulphide bonds (Fig. 2b and Supplementary Fig. 9) that are also present in CXCR4, one connecting the N terminus to the extracellular start of TM7 (Cys 30–Cys 277) and the other connecting the extracellular end of TM3 to ECL2 (Cys 110–Cys 187). These Cys pairs are highly conserved in the sequences of chemokine receptors and are important for ligand binding. Together, they have an important role in shaping the extracellular structure of the receptor, and also provide useful restraints for structure determination. The long ECL2 of CXCR1 forms a β-hairpin, the structure of which is constrained by the Cys 110–Cys 187 disulphide bond. A similar structure is observed in CXCR4 and many other GPCRs, despite a lack of amino acid sequence similarity in this region.

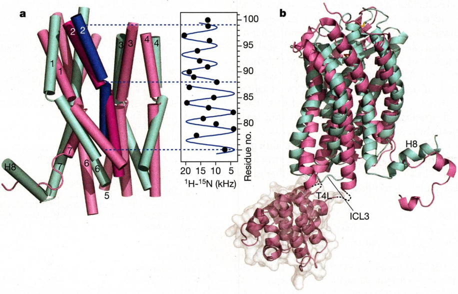


Figure 3 | Structural comparison of CXCR1 and CXCR4. CXCR1 (PDB accession 2LNL) is in cyan and CXCR4 (PDB accession 3ODU) is in pink. **a**, Comparison of transmembrane helices. TM2 (residues 74–101) of CXCR1 (blue) has a kink that changes helix direction at Phe 88. The kink is reflected in disruption of the dipolar wave near Phe 88. TM2 of CXCR4 (magenta) has a kink at the same location (Phe 87). **b**, Comparison of backbone structures. The third intracellular loop (ICL3) of CXCR4 is replaced by T4 lysozyme (T4L, molecular surface representation). The C terminus of CXCR1 forms a well-defined amphipathic helix (H8), whereas that of CXCR4 is only loosely helical.

However, the ECL2 β-hairpin of CXCR1 is less well defined than that of CXCR4, consistent with the presence of two Pro residues in CXCR1.

In both CXCR1 and CXCR4, charged residues are mainly located near the membrane-water interface (Supplementary Fig. 10), with negative charges clustered in the extracellular loops, where they can play a part in ligand binding and receptor activation. In addition, four charged residues, contributed by helices TM2 (Asp 85), TM3 (Lys 117) and TM7 (Asp 288 and Glu 291), form a polar cluster in the core of the helical bundle of CXCR1 that may have important consequences for ligand binding and receptor signal transduction. One of these residues (Asp 288) is not conserved in CXCR4 and may contribute to the differences in biological activities of the two chemokine receptors.

The intracellular loops of GPCRs are crucial for G-protein interactions[11]. Modification of ICL3 by insertion of T4 lysozyme between TM5 and TM6, rendered CXCR4 incapable of activating G proteins[28]. By contrast, unmodified CXCR1 is fully active with respect to both G-protein activation and chemokine binding, and its three intracellular loops are structurally well defined (Figs 2 and 3 and Supplementary Fig. 8). Notably, ICL3, which is important for CXCR1 coupling to G proteins and is involved in calcium mobilization, chemokine-mediated migration and cell adhesion, extends from Thr 228 to Gln 236, connecting helices TM5 and TM6, both of which are one turn shorter than the corresponding helices in CXCR4. ICL3 protrudes into the cytoplasm where it is available for G-protein binding. The sequence of ICL3 shares significant homology with other GPCRs; therefore, the ability to observe the structure of an intact GPCR provides an opportunity to propose structure-based mechanisms of G-protein binding and activation.

CXCR1 has been the subject of significant molecular modelling efforts aimed at understanding its interactions with small molecule inhibitors, including compounds active in reducing breast cancer metastasis[30]. The structure of CXCR1 determined in a lipid bilayer membrane should facilitate these studies. The solid-state NMR approach used for structure determination has several significant advantages. The protein resides in fully hydrated liquid crystalline phospholipid bilayers at physiological conditions of temperature and pH, and no detergents or non-native lipid phases are present. The protein sequence is unmodified, with no truncations, mutations or insertions of foreign proteins. The phospholipid composition can be varied, and other components, such as cholesterol, can be added. Other proteins or small molecules, including chemokines, G proteins, nanobodies and drugs can be added directly to the samples, enabling detection of any structural changes by direct spectroscopic and structural comparisons. NMR is adept at describing both overall and local protein dynamics. Measurements of DC values and isotropic chemical shifts to provide molecular orientation and dihedral angle restraints, combined with refinement in a membrane environment, facilitate structure determination by molecular fragment replacement. These can be supplemented with orientation restraints derived from rotationally averaged chemical shift anisotropy powder patterns to improve structural accuracy and precision further. Thus, we anticipate that this method will enable structure determination and structure-activity studies of other GPCRs, as well as a wide range of other membrane proteins under near-native conditions.

METHODS SUMMARY

The native sequence of human CXCR1 (residues 1–350) was expressed in *E. coli*, using M9 minimal media containing ^{15}N-labelled ammonium sulphate and $^{13}C_6$-glucose or 2-^{13}C-glycerol (Cambridge Isotope Laboratories), then purified by chromatography, refolded in DMPC proteoliposomes by detergent dialysis, and assayed for IL-8 ligand binding and G-protein activation[22,23]. Samples for NMR studies consisted of reconstituted CXCR1 proteoliposomes suspended in buffer, concentrated by ultracentrifugation and packed into an MAS rotor. An NMR sample contained 2.5 mg of CXCR1 and 12.5 mg of DMPC (1:300 protein:lipid molar ratio) in a 36-μl volume.

NMR experiments for resonance assignments and the measurements of structural restraints (Supplementary Table 2) were performed at 25 ± 1 °C on a 750 MHz Bruker Avance spectrometer equipped with a Bruker low-E ^{1}H/^{13}C/^{15}N triple-resonance 3.2-mm MAS probe. The rotor spinning rate was controlled to 11.111 ± 0.002 kHz. Orientation restraints were derived from measurements of ^{1}H-^{15}N and ^{1}H-^{13}Cα DC. Backbone dihedral angles were derived from measurements of isotropic chemical shifts.

Structure calculations were performed in three stages. First, a database of short molecular fragments was generated from the sequence of CXCR1 and the assigned isotropic chemical shifts using CS-Rosetta[20]. Second, the fragments were used to generate and refine 1,000 structural models using the implicit membrane potential of Rosetta[19,21] with the experimental restraints and the structure of rhodopsin (PDB accession 1F88)[25] as topology template. Convergence of the Rosetta calculations is shown in Supplementary Fig. 11. Third, the lowest energy structure was selected for further refinement against all experimental restraints, using simulated annealing in Xplor-NIH[26]. A total of 100 structures was calculated and the 10 lowest energy structures were accepted for the structural ensemble. NMR and structure refinement statistics are provided in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper.

Received 2 February; accepted 12 September 2012.

Published online 21 October 2012.

1. Holmes, W. E., Lee, J., Kuang, W. J., Rice, G. C. & Wood, W. I. Structure and functional expression of a human interleukin-8 receptor. *Science* **253,** 1278–1280 (1991).
2. Sallusto, F. & Baggiolini, M. Chemokines and leukocyte traffic. *Nature Immunol.* **9,** 949–952 (2008).
3. Waugh, D. J. & Wilson, C. The interleukin-8 pathway in cancer. *Clin. Cancer Res.* **14,** 6735–6741 (2008).
4. Rajagopal, S., Rajagopal, K. & Lefkowitz, R. J. Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nature Rev. Drug Discov.* **9,** 373–386 (2010).
5. Goncalves, J. A., Ahuja, S., Erfani, S., Eilers, M. & Smith, S. O. Structure and function of G protein-coupled receptors using NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **57,** 159–180 (2010).
6. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459,** 356–363 (2009).
7. Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33,** 17–27 (2012).
8. Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function. *Science* **318,** 1266–1273 (2007).
9. Rasmussen, S. G. *et al.* Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450,** 383–387 (2007).
10. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl Acad. Sci. USA* **93,** 14532–14535 (1996).
11. Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature. Rev. Mol. Cell Biol.* **9,** 60–71 (2008).
12. Crump, M. P. *et al.* Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *EMBO J.* **16,** 6996–7007 (1997).
13. Rajagopalan, L. & Rajarathnam, K. Ligand selectivity and affinity of chemokine receptor CXCR1. *J. Biol. Chem.* **279,** 30000–30008 (2004).
14. Park, S. H., Casagrande, F., Cho, L., Albrecht, L. & Opella, S. J. Interactions of interleukin-8 with the human chemokine receptor CXCR1 in phospholipid bilayers by NMR spectroscopy. *J. Mol. Biol.* **414,** 194–203 (2011).
15. Das, B. B. *et al.* Structure determination of a membrane protein in proteoliposomes. *J. Am. Chem. Soc.* **134,** 2047–2056 (2012).
16. McDermott, A. Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. *Annu. Rev. Biophys.* **38,** 385–403 (2009).
17. Opella, S. J. & Marassi, F. M. Structure determination of membrane proteins by NMR spectroscopy. *Chem. Rev.* **104,** 3587–3606 (2004).
18. Edidin, M. Rotational and translational diffusion in membranes. *Annu. Rev. Biophys. Bioeng.* **3,** 179–201 (1974).
19. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77,** 363–382 (2008).
20. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105,** 4685–4690 (2008).
21. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using Rosetta. *Proteins* **62,** 1010–1025 (2006).
22. Park, S. H. *et al.* High-resolution NMR spectroscopy of a GPCR in aligned bicelles. *J. Am. Chem. Soc.* **128,** 7402–7403 (2006).
23. Park, S. H. *et al.* Optimization of purification and refolding of the human chemokine receptor CXCR1 improves the stability of proteoliposomes for structure determination. *Biochim. Biophys. Acta* **1818,** 584–591 (2012).
24. Park, S. H. *et al.* Local and global dynamics of the G protein-coupled receptor CXCR1. *Biochemistry* **50,** 2371–2380 (2011).
25. Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289,** 739–745 (2000).
26. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160,** 65–73 (2003).

27. Krebs, A., Edwards, P. C., Villa, C., Li, J. & Schertler, G. F. The three-dimensional structure of bovine rhodopsin determined by electron cryomicroscopy. *J. Biol. Chem.* **278,** 50217–50225 (2003).
28. Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330,** 1066–1071 (2010).
29. Mesleh, M. F. *et al.* Dipolar waves map the structure and topology of helices in membrane proteins. *J. Am. Chem. Soc.* **125,** 8928–8935 (2003).
30. Ginestier, C. *et al.* CXCR1 blockade selectively targets human breast cancer stem cells *in vitro* and in xenografts. *J. Clin. Invest.* **120,** 485–497 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by grants R01EB005161, R01GM075877, R21GM94727, R21GM075917, P01AI074805 and P41EB002031 from the National Institutes of Health (NIH). Further support came from Cambridge Isotope Laboratories. F.C. was supported by fellowships from the Swiss National Science Foundation (PBBSP3-123151) and the Novartis Foundation.

Author Contributions S.J.O. designed the study. S.P. optimized CXCR1 purification, refolding and NMR sample preparation. B.B.D. performed the NMR experiments. H.J.N. assisted in NMR data analysis. H.K. developed initial protocols for CXCR1 purification, refolding and functional assays. K.M. assisted in the revision of these methods for NMR experiments. A.A.D. tested samples for their suitability for NMR experiments. F.C., M.C. and K.M. expressed and purified CXCR1. F.M.M. and Y.T. performed the structure calculations. S.J.O., S.P., B.B.D., F.M.M. and Y.T. prepared the figures and wrote the paper.

Author Information The atomic coordinates for residues 29–324 of CXCR1 and NMR restraints have been deposited in the Protein Data Bank (PDB) under accession 2LNL. Assigned NMR frequencies have been deposited in the Biological Magnetic Resonance Bank (BMRB) under accession 18170. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.O. (sopella@ucsd.edu).

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature11668

Corrigendum: Yeretssian *et al.* reply

Garabet Yeretssian, Ricardo G. Correa, Karine Doiron, Patrick Fitzgerald, Christopher P. Dillon, Douglas R. Green, John C. Reed & Maya Saleh

Nature **488,** E6–E8 (2012); doi:10.1038/nature11367

In our Reply to the Brief Communications Arising 'Is BID required for NOD signalling?' we made use of a figure generated by the authors of the Comment (*Nature* **488,** E4–E6 (2012); doi:10.1038/nature11366) as part of the review process, in the bottom panel of our Fig. 1a. We introduced two errors (by inadvertently removing five data points, two from the wild type and three from *Ripk2*$^{-/-}$) into this figure when we were asked to provide a higher-resolution version at the production stage. The corrected panel is shown below as Fig. 1, representing serum MCP-1 levels measured 4 h after MDP injection by multiplex, from wild type ($n = 10$), *Bid*$^{-/-}$ ($n = 9$) and *Ripk2*$^{-/-}$ ($n = 7$) mice.

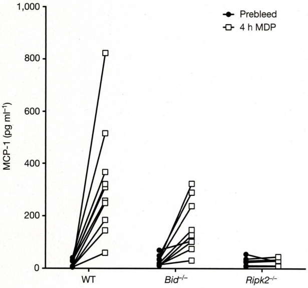


Figure 1 |

CORRIGENDUM

doi:10.1038/nature11694

Corrigendum: CD95 promotes tumour growth

Lina Chen, Sun-Mi Park, Alexei V. Tumanov, Annika Hau, Kenjiro Sawada, Christine Feig, Jerrold R. Turner, Yang-Xin Fu, Iris L. Romero, Ernst Lengyel & Marcus E. Peter

Nature **465,** 492–496 (2010); doi:10.1038/nature09075

In Fig. 1f of the original Letter, an incorrect actin blot was published: see the corrected panel in Fig. 1 of this Corrigendum. Also, in the original Supplementary Fig. 12c, some of the western blot data were either misinterpreted or raw data could not be located. We have now repeated the entire experiment: see Supplementary Information to this Corrigendum for the corrected Supplementary Fig. 12c. Although there are differences between the different experiments, the increase in phosphorylation of JNK and Jun was reproducible, confirming that stimulation of CD95 causes activation of JNK. All the conclusions of the original Letter are intact except for the data for the original Fig. 4f and g on the phosphorylation level of c-Jun and JNK in the livers of CD95-deficient mice, which have been corrected in two previous Corrigenda: *Nature* **471,** 254 (2011); doi:10.1038/nature09897 and *Nature* **475,** 254 (2011); doi:10.1038/nature10221. The results on the phosphorylation of JNK and Jun in mice injected with the murine CD95-specific agonistic antibody Jo2 under non-apoptotic conditions stand (both at the level of western blot and immunohistochemistry) and are not affected by the above changes; nor are any of the other figures. Further, the key findings of the Letter on the role of CD95 as a growth promoter in cell lines, in endometrioid, ovarian or liver cancer, and in liver regeneration are not affected by these corrections. For clarity, we now provide all the raw western blot data for the original figures and the corrected figures as Supplementary Information to this Corrigendum. A. Hadji and S. DeChant from Northwestern University generated data for the corrigenda. L.C. has declined to sign this Corrigendum.

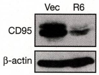


Figure 1 | This figure is the corrected inset to Fig. 1f. CD95-expression levels of HepG2 cells expressing scrambled control (Vec) or CD95shRNA6 (R6), as determined by western blot analysis.

Supplementary Information is available in the online version of the Corrigendum.

Product Focus: Hottest New Technologies

Spotlight!

2012's latest technologies

An impressive array of new technologies to support advancements in many different fields of research have emerged during 2012. This feature takes a look back at some of these hot new technologies.

Affymetrix provides high-quality solutions to advance science in basic and translational research. One major hurdle in advancing expression-based biomarker signatures has been the ability to generate high-quality data from RNA isolated from highly valuable formalin-fixed paraffin-embedded (FFPE) tissue sources. Affymetrix has launched the **Sensation Plus™ FFPE Reagent Kit**, which is capable of generating robust gene expression signatures from old (10+ year-old) FFPE samples starting with as little as 20 ng input RNA. The Company has also introduced **Axiom® miRNA Target Site Genotyping Arrays**, the only high-density genotyping tools for genome-wide evaluation of microRNA (miRNA) target sites.

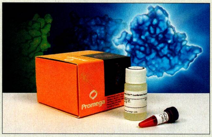
Transcriptome sequencing, or RNA-Seq, allows a global, unbiased view of gene expression changes, including novel transcripts, isoforms, and post-transcriptional modifications. Compared to gene expression microarrays, the approach can provide a wider dynamic range for detecting rare transcripts. "One barrier to adoption, however, is the perceived difficulty in comparing prior microarray data with RNA-Seq results. That isn't necessarily the case," says Wendell Jones, Ph.D., Vice President of Statistics and Bioinformatics at **Expression Analysis**. Jones' team at Expression Analysis developed the first method for translating output from the Company's **RNA-Seq services** into the existing **Affymetrix CEL** and other microarray file formats.

High-content peptide microarrays with custom and off-the-shelf contents for immunological research are provided by **PEPperPRINT**. The **PEPperCHIP® Peptide Microarrays** are synthesised with a patented laserprinter-based method directly on the chip. The benefits of this approach are a unique flexibility in terms of custom peptide content, a high spot density and reduced material consumption. Microarrays are provided on conventional object slides (containing 9,000 individual peptides) and other glass slide formats with up to 275,000 peptide spots. Assays can be performed by fluorescently labelled proteins or sandwich immunoassays.

Life Technologies has launched three next generation **GeneArt® Genetic Engineering Kits** that allow molecular and synthetic biologists unprecedented speed, flexibility, precision, and efficiency for the seamless cloning, assembly, and editing of genetic material. The **GeneArt® Seamless PLUS Cloning and Assembly Kit** allows investigators to complete complex assembly projects in days, rather than weeks with typically available methods. The **GeneArt® Seamless Cloning and Assembly Enzyme Mix** is the economical choice for creating constructs up to 13kb with the option for high-throughput assembly. The **GeneArt® Site-Directed Mutagenesis PLUS System** can be employed to introduce deletions, insertions and substitutions ranging from small to large fragment sizes and can facilitate single or multi-site mutagenesis.

Stem cell research, toxicology studies, and flow cytometry all rely on an accurate and consistent number of input cells **Bio-Rad Laboratories, Inc.** has launched its **TC20™ automated cell counter**, the next generation of the Company's TC10™ automated cell counter that provides accurate and reliable counts of live mammalian cells in 30 seconds. The TC20 system's improved lens and cell counting algorithm make it compatible with a broader range of cell sizes and types. This includes cells smaller than 10 microns and complex samples such as stem cells and primary cells derived from tissue or blood.

Ideal for pharmaceutical and clinical research laboratories, **Agilent Technologies Inc.** has introduced the **Agilent Automated Card Extraction LC/MS system**, a fully integrated instrument for the analysis of dried blood spots and other dried media. This system provides an automated workflow solution to improve productivity and streamline sample processing. Dried blood-spot analysis can be used to obtain detailed quantitative data about the compounds in animals and human subjects, using small volumes of blood (typically 10 to 20 µl). The stability of blood spots is often superior to frozen plasma or serum samples


Promega's new luciferase, NanoLuc™

The first commercially available kit to simultaneously interrogate the relationship between nucleosome occupancy, transcription factor binding and DNA methylation within the same DNA molecule is available from **Active Motif**. Traditional methods have only been able to look individually at DNA methylation or nucleosome positioning, without the ability to understand the relationship between the two. The **NOMe-Seq** (Nucleosome Occupancy and Methylome Sequencing) **assay** utilises fixed chromatin, which preserves nucleosome and protein binding locations. NOMe-Seq provides researchers with a tool to better understand nucleosome occupied or depleted chromatin in the context of methylation for its effects on gene regulation.

Promega has introduced a new luciferase, **NanoLuc™**, a smaller, brighter, and more versatile bioluminescent enzyme, enabling more advanced reporter assays with potential in biologically complex applications that require greater sensitivity. Developed with a novel substrate, furimazine, its unparalleled small size (a 19kDa monomer) allows for enhanced viral delivery and protein fusion applications, easy

Material compiled by College Hill

www.collegehill-lifesciences.com

CollegeHill

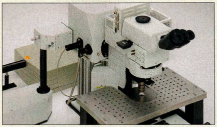
Product Focus: Hottest New Technologies

TC20™ automated cell counter from Bio-Rad Laboratories

secretion from cells, and uses where the gene or protein size is limiting. The enzyme is two orders of magnitude brighter than either firefly (Photinus pyralis) or Renilla reniformis luciferases, resulting in improved performance in hard to transfect cells.

Olympus has released the new **FV10MP-BXD-GAP photodetector** for enhanced sensitivity when performing multi-photon fluorescence imaging. The new non-descanned detector unit is ideal for capturing very faint fluorescence signals and for minimising the laser power required to generate a detectable signal, making it the best choice for advanced life science research. Exceptional sensitivity is achieved by coating the photoelectric surface of the detector with Gallium Arsenide Phosphide (GaAsP), which offers superior quantum efficiency compared to conventional photomultiplier tubes.

NanoZoomer-XR, the newest addition to the NanoZoomer product line for whole-slide scanning has been released by **Hamamatsu Photonics K.K.**. The new NanoZoomer-XR minimises workload and slide scanning time by automatically and continuously scanning up to 320 slides over two times faster than the NanoZoomer HT and RS scanners. It converts a 15 mm x 15 mm area on a glass slide into a 1.1-gigapixel colour image in as little as 30 seconds. The scanned slides are viewed smoothly and quickly through the newly developed viewer software (NDP.view2), and easily analysed using the IHC analysis software (NDP.analyze).



The FV10MP-BXD-GAP photodetector from Olympus.

NanoTemper Technologies has launched the **Monolith NT.LabelFree**, the first truly label-free and immobilisation-free instrument capable of measuring biomolecular interactions using Microscale Thermophoresis (MST). This technology measures the directed motion of molecules through a local temperature gradient detecting a dynamic range of 10nM to mM dissociation constants, while only consuming <4µl of sample. The NT.LabelFree can monitor the binding of single ions (40Da) or small molecules (300Da) to a target as well as the binding of ribosomes (2.5MDa). The NT.LabelFree is easy to handle and can generate data in 10 minutes for medium through-put purposes.

BMG LABTECH's innovative technology greatly enhances microplate applications and can improve scientific research. For example, the **SoPRano™ label-free technology** in a simple spectrometer-based absorbance reader will obtain protein-protein binding kinetics such as the association (k_a) and dissociation (k_d) rates, as well as EC/IC50 and K_D values. With SoPRano™ assays and BMG LABTECH's ultra-fast, full spectrum analysis using a CCD-spectrometer, label-free applications are now possible without purchasing expensive, dedicated SPR equipment. In addition, cell-based assays greatly benefit from the Company's innovative Direct Optic Bottom Reading design, which when used on the **PHERAstar FS** microplate reader achieves signal-to-blank ratios for specific fluorescent cell-based assays that are more than 300% higher than other microplate readers

Expanding on its innovative benchtop CoolBox technology, **BioCision LLC**, a leader in advanced temperature management of biomedical samples, has launched the **CoolBox™ XT workstations**. "Our new CoolBox XT technology addresses laboratory demands for extended cooling technology that provides the temperature of ice, without the intrinsic problems that come with its use," said Rolf Ehrhardt, CEO, BioCision. "The CoolBox XT workstation outperforms currently available systems based on its extended cooling and freezing times, independence from power, and unique size options."

Molecular Devices has released **version 2.5 of its MetaMorph® NX Microscopy Automation and Image Analysis Software**, which now includes modules for the study of neuronal morphology and fiber structures, support for targeted illumination devices, and a high-speed image acquisition mode. The Neurite Tracing Application Module simplifies the challenging, meticulous tasks associated with neuron research, while the Fiber Tracing Application Module quickly makes sense of complex, interwoven fiber structures.

A powerful simulation toolkit for neuroscientific experimentation has been released by **Intific, Inc.**. The **Real World®** with **NeuroBridge™ software platform** allows researchers to easily develop and explore complex immersive experiences in order to understand human cognition. It is a leap ahead in neuroscience software, providing engaging experiences as well as a mechanism to feed back real-time brain activity to the simulation experience. The NeuroBridge™ software enables correlation of simulation events with dynamic brain activity (EEG) and other physiological signals.

Companies mentioned in this Product Focus

Active Motif	www.activemotif.com
Affymetrix	www.affymetrix.com
Agilent Technologies	www.agilent.com
Bio-Rad Laboratories	www.bio-rad.com
BioCision	www.biocision.com
BMG LABTECH	www.bmglabtech.com
Expression Analysis	www.expressionanalysis.com
Hamamatsu Photonics	www.hamamatsu.com
Intific	www.intific.com
Life Technologies	www.invitrogen.com
Molecular Devices	www.moleculardevices.com
NanoTemper Technologies	www.nanotemper-technologies.com
Olympus	www.olympus.com
PEPperPRINT	www.pepperprint.com
Promega	www.promega.com

"This article was compiled by College Hill and submitted to Nature. It has not been written by or reviewed by the Nature editorial team and Nature takes no responsibility for the accuracy or otherwise of the information provided. Submit press releases for consideration to productfocus@nature.com with the topic in the subject line."

TURNING POINT Award provides crucial funding for glaciochemist **p.787**

SALARIES More men than women negotiate when pay is defined, says study **p.786**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

IMAGES.COM/CORBIS

COLUMN

When lab leaders take too much control

Two 'toxic' management models treat trainees poorly and leave labs at risk of misconduct, says **Charles Wood**.

Throughout her very active career, my graduate adviser, Mary Dallman, has insisted on having a table in the laboratory. I remember how Dallman, a neuroscientist at the University of California, San Francisco, would gather her graduate students and postdocs at that table to discuss data, design experiments and challenge the dominant paradigms of the time.

She had an office, of course, but the day-to-day business was done at that table, in the middle of the laboratory. She made sure never to have more than three graduate students and two postdocs at a time, and everyone participated. We critiqued each other, but were supportive if experiments didn't go as expected, or produced unexpected results. We knew that apparent failures might indicate even more exciting lines of enquiry.

Dallman taught me most of what I learned about how to run a lab. I don't have a central table (animal research means that some members of my group must work in a separate facility), but I keep my lab small, and I encourage students to foster discussions, question my ideas and update me on their progress, good and bad.

I have studied and observed how other labs are run worldwide, and have learned that good mentorship is not only a means of fostering capable, independent principal investigators, but also the best protection against misconduct. Just as there are great models of mentorship, there are also what I call toxic models, which can both discourage trainees and encourage misconduct. The 'executive' model involves a micromanaging principal investigator with a heavy hand; the 'competition' model endorses cut-throat and counter-productive rivalry among lab members (see 'Toxic teams'). I advise any student to avoid training in or collaborating with such labs.

SUCCESS AT ALL COSTS

In the executive model of lab management, the principal investigator demands that trainees meet his or her expectations, often with a specific goal in mind. In its most toxic form, that goal can include specific experimental outcomes — so a trainee is told to do this experiment and get this particular result. There is no room for disagreement, or for the trainee to overturn the investigator's paradigm.

I learned about the dangers of this model from one misconduct case in which a large, successful group partnered with other labs and set a very powerful, clear goal, but the principal investigator rarely spoke to students ▶

SALARIES

Pay negotiations studied

Women are slightly more likely than men to bargain for higher pay when a job advert indicates that salary is negotiable, a study finds. But men tend more than women to ask for more money when it is not made explicit that wages can be adjusted, says *Do Women Avoid Salary Negotiations? Evidence from a Large Scale Natural Field Experiment*, a working paper published on 15 November by the US National Bureau of Economic Research. Researchers placed job adverts for real administrative positions in nine US cities between November 2011 and February 2012, drawing almost 2,500 respondents. They found that 11% of men and 8% of women initiated salary negotiations when the salary was fixed, whereas 24% of women and 22% of men started discussions when it was negotiable. Study co-author John List, an economist at the University of Chicago in Illinois, suspects that the pattern is probably the same for scientific research positions. “Even if a job advert says the salary is not negotiable, women should negotiate — unless they want to stay a step behind,” he says.

FACULTY HIRING

Adjuncts lack support

Three-quarters of US academic institutions polled in a survey reported increasing their numbers of full-time non-tenure-track — or adjunct — faculty members in the past decade. More than one-third have “significantly increased” hiring of part-time adjuncts in the same period, finds *Values, Practices and Faculty Hiring Decisions of Academic Leaders*, a study that will be published in early 2013 in the journal *Liberal Education*. However, the paper reports that just 58% of the 157 responding institutions offer structured mentoring to full-time adjunct faculty members, and only 42% provide professional-development opportunities such as workshops on writing grant applications and managing grant budgets. Adrianna Kezar, co-author of the study and an associate professor of higher education at the University of Southern California in Los Angeles, says that early-career researchers interviewing for adjunct positions should negotiate for professional-development support and mentoring, which help to make candidates indispensable at universities and may confer an advantage on those who attempt to make the jump to a tenure-track post.

► or collaborators and never observed primary data being produced. Not surprisingly, a junior faculty member in a collaborating group eventually manipulated data to meet the goal, and no one realized for years. Of course, some scientists are more tempted to cheat than others whatever the management style, but cheating can be prevented by an involved principal investigator who is open to whatever data are produced. It can be disastrous to pressurize lab members — subtly or obviously — to produce only data that support the principal investigator’s hypothesis, particularly when fellow trainees lose their jobs for producing data that don’t meet the requirements (a penalty that, for foreign students, can mean deportation). What’s more, in such a high-pressure, isolated environment, principal investigators and other collaborators often fail to teach students important lessons, such as how their portions of the project fit in with the larger goal. The lessons that do get learned are negative: competition over collaboration and conformity over creativity. Once they leave the lab, students either drop out of science or go on to run their own labs on the same model.

HEAD TO HEAD

In the second toxic style of mentorship, the competition model, principal investigators give two or more trainees the same goal. The implication is that the one who completes the task first — or, more dangerously, the one who generates the data that conform best to the preconceived outcome — is the winner.

Often, the competition is not as obvious as giving trainees the same project, but it still means that they compete against each other, perhaps for first-authorship of a paper or credit for collaborative work. If a trainee does not win the prize, he or she will face much poorer career prospects. This creates the perfect motivation to cross the line and fabricate data, and makes everyone in the lab unhappy and suspicious of one another. Students are treated purely as labour, with no regard to their education. If they do win, they often go on to become bad mentors themselves.

WARNING SIGNS

Toxic teams

The ‘executive’ and ‘competitive’ lab-leadership models can create poor work environments with undue pressure to perform, says Charles Wood. Here are some signs to look for:

Executive

- Lab members spend little time with the principal investigator.
- The lab is busy, and focuses only on high-impact papers and large grants.
- Research goals are clearly established.
- Penalties for unmet goals are strong.
- There is little cross-talk with collaborators.

Competitive

- One goal is assigned to several people.
- The lab is large, with many trainees.
- Lab members are unhappy and hyper-competitive.
- Socializing is rare.

THE ANTIDOTE

I have seen many labs run on these toxic models, although not always in such extreme forms. Principal investigators often opt for shades of one or both. To avoid making the same mistakes myself, I am constantly examining my own relationships with my trainees, being sure to limit the number I take on. At the moment, my lab consists of only four graduate students, and no postdocs. Certainly there are mentors who can successfully advise many students: a large group does not by itself lead to a toxic model of mentorship. And different scientific disciplines need different amounts and types of mentorship. But increasing the number of trainees can dilute the mentoring experience. The balance of time and commitment will certainly be different for each faculty mentor.

I often tell my students that ‘academic scientist’ is the best possible job. The joy that comes from individual mentorship, from discovery, from intellectual and practical challenges, is unique to this environment. Trainees should experience the fun of science, teaching and learning, not the toxic environment of a dictatorial enterprise. And they should be able to question approaches and orthodoxy. I always encourage dissent and enable trainees to question a working hypothesis.

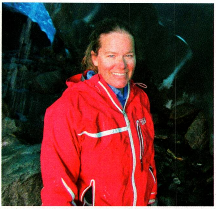
All successful models of mentorship have a common thread: the mentors commit to their trainees. Losing focus on the student or postdoc violates the most basic premise of mentorship — that trainees are there to be trained. Viewing trainees as merely cheap labour leads to toxic mentorship.

My advice to students and postdocs is to choose your adviser well. Pick an open mentor who has a good track record with students. If your principal investigator starts to exhibit toxic behaviour, address this with him or her. If you find yourself in a truly toxic environment, seek guidance from a graduate coordinator, assistant dean or other authority figure who oversees the pre- or postdoctoral training programmes — and ask for help in finding another mentor. No fledgling scientist has time to waste on a toxic situation. ■

Charles Wood *is chair of the department of physiology and functional genomics at the University of Florida in Gainesville.*

TURNING POINT
Sarah Aciego

Sarah Aciego, a glaciochemist who has pioneered isotope dating of ice cores at the University of Michigan in Ann Arbor, won a five-year US$875,000 grant from the David and Lucile Packard Foundation in Los Altos, California, on 15 October.



How did events at university help to set you on your current path?
I ended up as a geology major based in the engineering school, and therefore took chemistry, physics and calculus, learning skills that facilitated my exposure to isotope work. I didn't have much money, so I needed a work–study job. I distributed flyers advertising my calculus and computer-programming skills around the department, and was hired to write programs to process isotope data looking at dust inputs and silica cycling in soil ecosystems.

How did your work with isotopes continue?
I applied to a number of graduate schools, and came to work with an isotope chemist who was using radiogenic isotopes to examine volcanic processes. But a change to the PhD requirements the year I arrived meant that I had to present and defend two possible thesis projects during qualifying exams, so I had to find a second adviser and develop an alternate project, working with a glaciologist who was using stable isotopes to date ice sheets in Antarctica. I was so interested in both topics that I did two PhD projects.

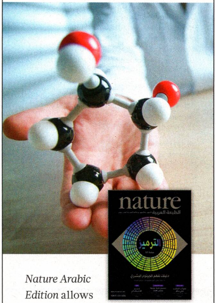
How did that affect your opportunities?
There is a pretty big divide between the radio- and stable-isotope research communities. But my combined efforts made me realize where my expertise could make a difference. Dating the deepest ice had proved nearly impossible using conventional methods, so I wanted to use a radiometric isotope technique to date ice using dust. I did that for my postdoc at the Swiss Federal Institute of Technology in Zurich, where a talented mass spectrometrist and engineer, Heinrich Baur, developed both a device called a nanoscale that measures nanogram variations in mass, and a technique to measure the weight of dust and the weight of the gas absorbed on the dust.

You started at Michigan in 2010. Were you surprised by the generous support you received despite the economic downturn?
Yes. I interviewed at six places over two years and didn't think I had a chance in hell of anyone buying me the equipment I needed. Luckily, the University of Michigan realized the value of the machines and clean lab facilities necessary to run the kind of experiments I wanted to do. And the US National Science Foundation partially funded me to get one of the $200,000 nanoscales developed in Zurich. In return, I try to serve as a resource to the field and to my students, and try to bring what I do into the classroom and to the public.

What does the Packard award allow you to do?
The Packard grant is huge because I can redirect some of the money, which is unrestricted, to my projects that are having a harder time getting funded. It is also helpful because the fieldwork for this research is so expensive. I have proved that my methods work in Antarctica, but this funding allows me to collect samples across Greenland to see if it works there as well. If it does, it will provide a much-needed comparison to clarify the impacts of climate change on sea-level rise. We will learn something — although it may not be what we set out to find.

Do you network naturally or out of necessity?
Necessity. I have learned that you have to put yourself out there and create a broad base of support from your community to garner recognition. A lot of those interactions happen during social events after conference talks. One of the nicest things that I read in reviews of my proposals, even those that have been rejected, is that I'm a rising star in the field and know what I'm doing. These comments are from people I haven't worked with directly — they have seen me in some public forum. I impress on my students that we go to meetings to advertise not only the science, but also our capabilities. ■

INTERVIEW BY VIRGINIA GEWIN


nature
الطبعة العربية

High impact science
for the Arabic community

Nature Arabic Edition allows Arabic speakers throughout the world to access top quality science news and comment from *Nature*, as well as summaries of all the research papers from the leading multidisciplinary journal. Content is freely available online every week, with limited monthly copies available in print.

Access *Nature Arabic Edition* online and apply for your free subscription

arabicedition.nature.com

In partnership with:

مدينة الملك عبدالعزيز للعلوم والتقنية KACST

nature publishing group npg

naturejobs

European Head Office, London naturejobs@nature.com
The Macmillan Building, 4 Crinan Street, London NI 9XW, UK
Tel: +44 (0) 20 7843 4961 Fax: +44 (0) 20 7843 4996

Global Head:
Nils Moeller,+44 (0) 20 7843 4953 n.moeller@nature.com
Southwest UK/RoW:
Emma Hare, +44 (0) 20 7843 4944 e.hare@nature.com
Northeast UK/Ireland:
Matthew Ward, +44 (0) 20 7014 4059 m.ward@nature.com
France/Switzerland/Belgium:
Julie Aramburu, +44 (0) 20 7843 4994 j.aramburu@nature.com
Scandinavia/Spain/Portugal/Italy:
Megan Ball, +44 (0) 20 7843 4973 m.ball@nature.com
North Germany/The Netherlands/Eastern Europe:
Kristin Doelling, +44 (0) 20 7843 4970 k.doelling@nature.com
South Germany/Austria:
Maria Kubalova, +44 (0) 20 7014 4084 m.kubalova@nature.com
Natureevents:
Matthew Clare, +44 (0) 20 7014 4015 m.clare@nature.com
Business Development Manager, Middle East:
Jon Giuliani, +44 (0) 20 7843 4922 j.giuliani@nature.com

USA Head Office, New York naturejobs@us.nature.com
75 Varick Street, 9th Floor, New York, NY 10013-1917
Tel: +1 800 989 7718 Fax: +1 800 989 7103

US Sales Manager:
Ken Finnegan, +1 212 726 9248 k.finnegan@us.nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2–37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751, Fax: +81 3 3267 8746

Asia-Pacific Sales & Business Development Manager:
Yuki Fujiwara, +81-3-3267-8765 y.fujiwara@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To, +852 2811 7191 g.to@natureasia.com

India
3a, 4th Floor, DLF Corporate Park, Gurgaon, 122 002, India

Advertisement Sales:
Vikas Chawla, +91 1242881057 v.chawla@nature.com

UNIVERSITY OF LEEDS

Faculty of Mathematics and Physical Sciences
School of Chemistry

Chair in Organic Soft Materials

The University of Leeds is seeking to appoint a Professor of Organic Soft Materials who will undertake internationally leading research in the development of new methodologies for the synthesis and characterisation of functional organic materials, which may include structures deriving from directed-assembly, with potential applications in a variety of sectors, including energy, high value chemicals, healthcare, electronics and nano-medicine.

We seek candidates who will be able to evidence international academic links and effective engagement with and influence on international research agendas. You will provide academic leadership both by personal research and by the encouragement and stimulation of colleagues. You will inspire students through research-led teaching, which will translate your excellence in research and scholarship into learning opportunities for students.

Applications from women, who are currently under-represented in this area, will be particularly welcome. However, the final recruitment decision will be based solely on merit.

The salary, which is negotiable, will be within the Professorial range – minimum £59,302 p.a.

Preliminary enquiries about the post should be directed to Professor Dwayne Heard, Head of the School of Chemistry: e-mail: d.e.heard@leeds.ac.uk, telephone: +44 (0)113 343 6471.

Closing date: Wednesday 2nd January 2013

Interview date: Thursday 28th February 2013

For more information and to apply, please visit http://jobs.leeds.ac.uk/ and complete the Vacancy Search Criteria. Job Ref: MAPCH0049

We welcome applications from all sections of the community. All information is available in alternative formats please contact +44 (0)113 343 4146.

U231241R

WORKING TOWARDS EQUALITY AND DIVERSITY www.leeds.ac.uk

Quantitative Biology Cluster Hire
University of Maryland, College Park

College of Computer, Mathematical and Natural Sciences
College of Agriculture and Natural Resources
A. James Clark School of Engineering
School of Public Policy

The University of Maryland, College Park seeks multiple tenured/tenure track faculty in the area of Quantitative Biology. The University counts among its researchers leaders at the interface of biological and physical sciences, and includes historically strong, world-recognized research programs on non-linear dynamics and statistical thermodynamics. With this initiative, the campus plans to significantly expand this expertise to create an outstanding interdisciplinary program harnessing the modern tools of biology, physics and engineering. We invite exceptional scientists with strong backgrounds in biology, physics, mathematics, technology development and/or entrepreneurship, whose research is advancing the frontiers of knowledge and contributing to a quantitative and integrated understanding of life.

The appointments may be made at the Assistant, Associate or Full Professor level, commensurate with qualifications and experience. Outstanding research facilities and a generous start-up package will be provided. The successful candidates will be expected to maintain cutting-edge externally funded research programs that synergize with existing core groups on campus, and to participate in undergraduate and graduate teaching. Applicants must have a doctorate degree, an outstanding publication record and a commitment to excellence in teaching.

The University of Maryland, College Park is the flagship campus of the University System of Maryland and one of the most rapidly advancing public research universities in the country. The campus is a highly interactive environment including various Centers and Institutes, such as those formed through partnership and collaboration with the National Institutes for Standards and Technology (NIST), the U.S. Food and Drug Administration (FDA) and the National Institutes of Health (NIH). Close proximity to Washington DC, Baltimore and the Maryland Biotechnology Corridor facilitates interactions with an extraordinary range of major research institutions such as NIST, NIH, NSF, DOE, NASA, FDA, USDA, NRL, ARL and HHMI-Janelia Farm, in addition to providing a rich cultural environment.

Applicants should apply electronically through Jobs@UMD at the following URL: http://jobs.umd.edu/applicants/Central?quickFind=57836

Applicants should submit the following: (1) cover letter, (2) curriculum vitae, (3) summary of research plans (maximum two pages), (4) one-page teaching philosophy statement and (5) the contact information with email of at least three references. Complete applications should be received by February 15, 2013, but will be accepted until the positions are filled.

The University of Maryland is an Affirmative Action/Equal Opportunity Employer.
Women and members of underrepresented groups are especially encouraged to apply.

NW231160R

UNIVERSITY OF MARYLAND

ARE YOU AS AMBITIOUS AS WE ARE?

The University of Exeter, a member of the Russell Group and firmly positioned in the UK Top 10 and the global Top 200, is embarking on a major expansion of its Medical School in both clinically-related and basic science research as part of an ambitious growth plan.

BASIC SCIENCE

In basic science, the Medical School is part of a multi-million pound interdisciplinary initiative with other Colleges of the University to investigate cellular dynamics, to develop a quantitative understanding of cell behaviour and the underlying biochemical and biophysical properties of cells and to determine how defects in cellular dynamics cause disease.

We are seeking to appoint world-leading researchers in the areas of Molecular Cell Biology, Genetics, Epigenetics, Systems Medicine, preferably with interests related to neuroscience, diabetes, vascular biology, ageing, and/or environmental challenge.

Either non-clinical or clinically qualified individuals may apply for the following basic science posts:

- Professors of Functional Cell Biology
- Professors of Systems Medicine
- Professors of Genetics/Epigenetics

CLINICAL AND HEALTH SERVICES RESEARCH

In clinical and health services research, the Medical School's ambitions focus on improving patient care, patient stratification, and health and wellbeing of the population. There are highly successful collaborations with health researchers in other departments of the University, particularly the world-leading Mood Disorders Centre, the Business School and in exercise science.

We are seeking to appoint world-leading researchers with expertise in clinical trials, health services or applied health research, experimental medicine investigating the mechanisms underlying disease or responses to treatments, diagnostics, innovative approaches to health data usage, systems medicine, or behavioural change.

The clinical research posts available are:

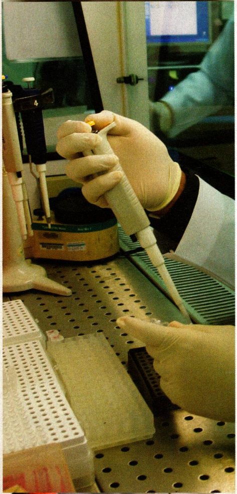
- Professors of Clinical or Health Research or Clinical Trials
- Professors of Complex Health Interventions
- Professors of Experimental Medicine
- The Gillings Chair in Applied Health Statistics
- Professors of Diagnostics

More detail on the School's research strategy and the posts can be found at **www.exeter.ac.uk/medicine/posts**

UNIVERSITY OF EXETER | MEDICAL SCHOOL

THE SUNDAY TIMES
UNIVERSITY OF THE YEAR 2012-13
UNIVERSITY OF EXETER

U231227R



VIVARIUM

RESEARCH FACILITY

The state-of-the-art animal Research Facilities at King Abdullah International Medical Research Center are designed according to international standards to ensure the safety and humane care of research animals. Three vivariums will be inaugurated this summer in Riyadh, Dammam and Al Hassa, respectively. The animal Research Vivarium in Riyadh covers a vast 7500 square meters and offers a unique research environment that combines housing, laboratories, imaging and surgery suites located in the same area. It comprises pathogen-free housing areas equipped with the newest individualized environmental cages that can hold up to 60,000 rodents. Likewise, automated animal housing and hygienic equipments are provided for rabbits, dogs, and sheep to warrant the highest possible care. Unique to the Gulf region, the facility also accommodates aquatic animals such as Zebrafish and Xenopus. The Vivarium Research Facility provides a complete range of equipment for recombinant DNA research, and Transgenesis, together with the latest generation of Optical Imaging, Positron Emission Tomography and Magnetic Resonance Imaging. Each vivarium is staffed with highly qualified veterinarians to ensure proper handling, care and use of animals in scientific research and education in accordance with the highest scientific, humane and ethical principles. Further, each Vivarium Research Facility is open for international accreditation and regular monitoring by the Institutional Animal Care and Use Committee.

- Saudi Arabia's Population : 28.08 million
- The largest spender on healthcare in the GCC with investments totaling $24.2bn last year, according to Frost & Sullivan
- Government plans to spend $373 billion between 2010 and 2014 on social development and infrastructure projects to advance economic development
- Set to be 6th richest economy in the World in terms of GDP per capita by 2050, according to a forecast by Citibank
- The largest free economic market in the Middle East, holding 25% of the total gross national Arab income

King Abdullah International Medical Research Center

Today's Question. Tomorrow's Answer

We are an international translational health science research institution that believes in creating a knowledge-based society through our channel of extraordinary human talent.

---

www.kaimrc.med.sa
research1@ngha.med.sa

RW228607R

南方科技大学

SOUTH UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

Faculty Positions at South University of Science and Technology of China(SUSTC)

SUSTC is a public institution funded by the municipality of Shenzhen, a special economic zone city in southern China. The University receives full financial support from the municipality. Set on five hundred acres of wooded landscape in the picturesque Nanshan (South Mountain) area, SUSTC offers an idyllic environment suitable for learning and scholarship. SUSTC promotes academic freedom, engages in basic and problem-solving research of lasting impact to benefit society and mankind, and strives to become one of the best research universities in the world.

Academic Disciplines:

SUSTC has a broad range of academic disciplines, covering natural science, engineering and some newly emerging inter-disciplines. All disciplines are open to applicants. The university especially encourages research that requires a multi-disciplinary and non-traditional approach. Successful applicants are expected to lead or establish new research directions at SUSTC.

Positions:

SUSTC invites applications and nominations from outstanding scholars of all nationalities for faculty positions at all ranks with full-time arrangements. Full-time positions include leading professor, chair professor, professor, associate professor, assistant professor, and PostDoc.

Qualifications & Responsibilities:

Applicants should have a doctoral degree in a relevant research field with outstanding communication skills and a proven record of research excellence.

Responsibilities include teaching undergraduate and graduate courses, developing and sustaining externally funded research programs, graduate student supervision, and professional/institutional services.

Recruiting Positions at:

- Leading Professor
  Applicants should have a Ph.D. degree with extensive research and teaching experience. Applicants should be internationally recognized as holding a leading position in important areas of research.
- Chair Professor
  Applicants should have a Ph.D. degree with extensive research and teaching experience. Applicants should possess a large number of internationally recognized research accomplishments.
- Professor
  Applicants should have a Ph.D. degree with extensive research and teaching experience. Applicants should have made a number of important contributions in research.
- Associate Professor
  Applicants should have a Ph.D. degree. Applicants should have made a few important contributions in research.
- Assistant Professor
  Applicants should have a Ph.D. degree. Applicants should possess evidence that demonstrates the ability in performing high quality research.

Salaries and Benefits:

The University offers competitive salaries and start-up packages. Benefits include housing, medical insurances, pensions, etc.

Contact Us:

Qualified applicants are invited to send a job intention letter and a curriculum vitae to: **hiring@sustc.edu.cn.**

Mailing Address:

Human Resources Department,
South University of Science and Technology of China
No.1088, Xueyuan Blvd, Xili, Nanshan, Shenzhen, China Zip Code: 518055

Telephone: +86-755- 86249985, 86249969

Relevant Links:

South University of Science and Technology of China: http://www.sustc.edu.cn/

JP221489R

---

the Beatson
institute for cancer research

Postdoctoral Research Scientist

**Starting Salary from £28,500 to £38,000
(Depending on experience)**

**Initial 3 year fixed term contract
(with possibility of up to 2 year extension)**

The Beatson Institute in Glasgow is one of Europe's leading cancer research centres. It is core funded by Cancer Research UK and supports cutting edge research into the molecular mechanisms of cancer development. The Institute provides an outstanding research environment, underpinned by state-of-the-art core services and advanced technologies with special emphasis on imaging and proteomics.

Control of the actin regulators WASP, SCAR and WASH in Dictyostelium
Prof Robert Insall

This Cancer Research UK funded project will investigate the way actin polymerization is controlled in cell migration and vesicle traffic. The Insall laboratory works on multiple aspects of actin, cell movement and chemotaxis. Most work will use Dictyostelium genetics, though new discoveries will be confirmed in cultured mammalian cells and in vivo models. The work will focus on activation of SCAR, WASP and WASH by dephosphorylation and other covalent modifications (see Ura et al., Curr Bio 22:553), and inactivation by proteolysis and interactions with actin binding proteins (Carnell et al., JCB 193:831-). Techniques involve culture of Dictyostelium, creation of complex, multiply knocked-out strains, generation and transfection of tagged markers, and high-level microscopy on multiple platforms.

A relevant PhD is essential. Experience in Dictyostelium and some of microscopy, cell biology, molecular biology and genetics, and a good record of quality publications would be an advantage. You will have good knowledge of the literature, enthusiasm and flexibility to work on a wide range of techniques as needed.

Cancer Invasion and Metastasis
Prof. Laura M. Machesky

The actin cytoskeleton drives invasive protrusion of cancer cells and organises adhesive and matrix degrading assemblies at leading edges. We have recently identified the actin nucleating protein N-WASP as a key protein involved in the assembly of invasive actin networks (see Yu et al., JCB 2012). We are currently seeking a postdoctoral fellow to further investigate the mechanisms by which N-WASP and other partner proteins organise invasion and generate actin networks that capture the matrix metalloprotease MT1-MMP, integrins and other receptors. The project also includes investigation of the role of actin nucleation in metastasis of melanoma in vivo.

You must have a relevant Ph.D. in molecular cell biology or murine genetics and experience with imaging and microscopy would be an advantage. You must be enthusiastic and motivated to work both with in vivo and cellular models.

Applications for the above posts should be sent with CV and names of three referees to Prof Robert Insall (r.insall@beatson.gla.ac.uk) or Prof Laura Machesky (l.machesky@beatson.gla.ac.uk) at The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow, G61 1BD, UK.

Closing date for applications: 1 January 2013

Scottish charity no SC006106

CANCER RESEARCH UK

U231239R

“Thank you for assisting in advertising the 6x ‘Research Associate’ positions on the Naturejobs website. I was advised thoroughly on the inventory that would be the best for our organisation and the web inventory worked. We received a huge response and will contact you for future advertisements.”

**Nanda Rahman,
Bio Nano Consulting, UK**

naturejobs

Translational Neuroscience Head of Preclinical Research

Boston Children's Hospital and Harvard Medical School

We seek an outstanding Head of Preclinical Research (PhD or MD, PhD) with industry experience to develop a vigorous, integrated translational neuroscience research program in an academic setting. Candidates should have 5-10 years preclinical development experience.

The successful candidate will have a leadership role in a new Translational Neuroscience Center, which resides within the very strong Program in Neuroscience and Neuroscience-related clinical departments and divisions at Boston Children's Hospital. The Center encompasses research and development as well as clinical trial activities and the role of the Head of Preclinical Research will be to facilitate and manage translational projects for the development of new therapeutics, biomarkers and diagnostics.

Specific job responsibilities to include:

1) Prioritizing which basic science discoveries in the neurosciences at Boston Children's Hospital are most suitable for translation.
2) Overseeing and encouraging cross-collaborations within the participating Cores of the Translational Neuroscience Center and amongst relevant departments at Boston Children's Hospital.
3) Identifying and running co-development partnerships with foundations, the NIH, and industry, to push individual translational projects forward.
4) Helping faculty members design studies to validate the targets/pathways for specific disease indications, develop suitable screening assays and run/analyze small scale screens to identify tool compounds, access high throughput screening and compound libraries for detecting new chemical entity hits, obtain medicinal chemistry advice and compound synthesis for lead optimization, test compounds in human neurons derived from patient iPSC lines, conduct preclinical pharmacological studies in relevant disease models for target exposure and engagement, efficacy, tolerability and pK, test for target specificity, rule out early safety issues in chemical series, run formal ADMET studies and perform pK/pD modeling.

The Head of Preclinical Research will hold both Children's Hospital and Harvard Medical School faculty appointments.

Deadline for applications is January 31, 2013. To apply, a current CV, description of drug development experience and research interests, and three to five reference letters and contact information should be sent to:

Translational Neuroscience Search Committee
Attn: Mustafa Sahin
Boston Children's Hospital, 300 Longwood Avenue, CLS13074, Boston, MA 02115
Maliha.Chaudhry@childrens.harvard.edu

Harvard Medical School and Boston Children's Hospital are equal opportunity/affirmative action employers with strong institutional commitments to diversity in their faculty. Women and minority candidates are particularly encouraged to apply.

NW231017R

BLUE WATERS

Tackling Some of Society's Biggest Scientific Challenges?

Think Illinois and the Blue Waters Supercomputer.

The University of Illinois at Urbana-Champaign's National Center for Supercomputing Applications is launching Blue Waters, one of the world's most powerful supercomputers with a peak performance of more than 11.5 petaflops. Building on more than a half-century of high-performance computing leadership at Illinois, Blue Waters is supported by the National Science Foundation.

Scientists around the country will use Blue Waters to tackle some of society's biggest challenges—understanding the building blocks of the universe, predictive geophysics, fundamentals of combustion, energy conversion and distribution, nanotechnology for electronics and medical devices, biomedical imaging and biophysical modeling, and the discovery and design of materials.

Part of Blue Waters could be yours. The University of Illinois has openings for several Blue Waters Professors, available to prospective faculty at any rank, who will receive substantial allocations on the supercomputer and expedited access to the system.

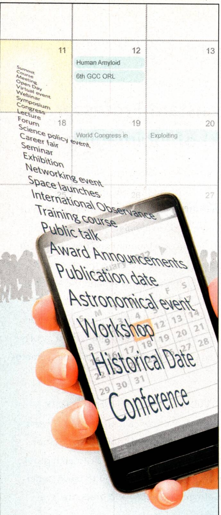
As part of the Visioning Future Excellence at Illinois initiative, Blue Waters faculty appointments will be made to individual departments within the College of Engineering.

For details on how to apply, go to: engineering.illinois.edu/bluewaters

Illinois is an Affirmative Action/Equal Opportunity Employer and welcomes individuals with diverse backgrounds, experiences, and ideas who embrace and value diversity and inclusivity. (www.inclusiveillinois.illinois.edu)

ILLINOIS

NW230553R



DISCOVER SCIENCE EVENTS — SYNC YOUR CALENDAR TODAY.

CALENDAR VIEW
NOW AVAILABLE AT
NATUREEVENTS.COM

- It's easy to sync our events database with your online or mobile calendar
- Find important dates in the science calendar, not just conferences

nature publishing group npg

UNIVERSITÄTSMEDIZIN
GÖTTINGEN **UMG**

The Department of General and Visceral Surgery at the University Medical Center Göttingen invites applications for a

Professorship

in Translational Research in Solid Tumors

(salary level W3 tenure track)

The professorship will initially be available for 5 years with the option of a permanent position.

The applicant should be able to represent the Department of General and Visceral Surgery with an emphasis on translational research in solid tumors both in research and teaching. The applicant's primary research interest should focus on molecular and functional analyses to study the development, progression, and treatment resistance of solid tumors.

We are looking for an integrative personality with outstanding scientific achievements and substantial external funding. The successful applicant will be expected to integrate into the scientific focus of the Department of General and Visceral Surgery and the Clinical Research Unit 179 (KFO 179), and to cooperate with existing research networks of the Göttingen Comprehensive Cancer Center and the Göttingen Research Campus. Initiation of new networks (Sonderforschungsbereich) is expected.

Further qualifications required for the position include a "Habilitation" or its equivalent in research and university teaching. The applicant should have a track record of successful applications for third party research funding and experience in coordinating research networks. Existing international collaborations are a prerequisite.

Further requirements for the appointment of university professors are laid out in § 25 of the Higher Education Act of the State of Lower Saxony dated Feb. 26, 2007 and modified by article 1 of June 10th, 2010 (Nds. GVBl. p. 242). Details will be provided upon request. The University of Göttingen reserves the right of appointment.

Applications from abroad are highly welcome. The University of Göttingen strives to increase the number of women on its scientific staff and therefore particularly encourages qualified women to apply for this position. Disabled persons of equal qualification will also be given preference.

Please submit your application online, including a curriculum vitae, a description of your scientific career, teaching activity, a list of publications, and a detailed research proposal, within 3 weeks using the following link:
http://www.med.uni-goettingen.de/de/content/ueberuns/berufungen_onlinebewerbung.html

For further information, please contact:
Berufungen@med.uni-goettingen.de

W231041R

"We recently placed an ad with naturejobs.com and the response was huge. We have now found the right candidate and will offer them the job"

Dr. Birgit Luber
Technische Universität
München, Germany

naturejobs

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Universität Zürich

Professor / Assistant Professor (Tenure Track) of Bioimaging

The Department of Information Technology and Electrical Engineering (www.ee.ethz.ch) at ETH Zurich and the Faculty of Medicine of the University of Zurich invite applications for a professorship or assistant professorship (tenure track) in Bioimaging. The open position is designated to focus on the development and use of imaging technology for biomedical applications, including the study and modeling of human physiology. It is aimed to expand the existing Bioimaging programme, which comprises magnetic resonance, molecular imaging, x-ray microscopy, and medical image analysis. The successful candidate's research may complement current efforts in these areas or focus on an additional technology, including emerging approaches.

Candidates should have a PhD degree and an excellent track record in Electrical Engineering, Biomedical Engineering, Physics, Biophysics, Computer Science, or related disciplines. In addition, commitment to teaching and the ability to lead a research group are required. The new professor will be expected to teach undergraduate level courses (German or English) and graduate level courses (English).

Your application should include your curriculum vitae, a list of publications and statements of your future teaching and research activities. The letter of application should be addressed **to the President of ETH Zurich, Prof. Dr. Ralph Eichler. The closing date for applications is 28 February 2013.** ETH Zurich is an equal opportunity and affirmative action employer. In order to increase the number of women in leading academic positions, we specifically encourage women to apply. ETH Zurich is further responsive to the needs of dual career couples and qualifies as a family friendly employer. **Please apply online at www.facultyaffairs.ethz.ch.**

W231215R

Change your environment. Find jobs where you'll make a difference
naturejobs

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

The Faculty of Medicine at Heinrich Heine University Düsseldorf invites applications for the permanent position of

Professor (W3) in Virology

The Heinrich-Heine-University searches for applicants who have an international reputation in the field of virology. The candidate should represent virology in university research, teaching and with given qualification in clinical diagnostics. The professorship is connected with the directorship of the Institute of Virology.

Besides a university education in the fields of human medicine or natural sciences, excellent scientific achievements, a successful international career, and a proven track record in the competitive acquisition of third-party funding are required. Applicants with academic studies in medicine should have a specialized training / residency in microbiology, virology and epidemiology of infections. The appointee has to be proficient in the laboratory diagnostics of viral diseases and has to strengthen the priority areas of clinical infectiology. The appointee is expected to foster international research at Heinrich Heine University Düsseldorf and contribute to its cooperative research projects (Collaborative Research Centres: 974 "Communication and Systemic Impact in Liver Damage and Regeneration"; Clinical Research Group: 217 "Hepatobiliary Transport and Liver Disease"; DFG Research Unit: 729 "Anti-infectious Effector Programmes: Signals and Mediators"; Research Training Groups: 1033 "Molecular Targets of the Aging Process and Strategies for the Prevention of Aging" and 1045 "Modulation of Host Cell Functions for the Therapy of Viral and Bacterial Infections").

Extensive experience in the management of departmental organizational units and in personnel management is necessary. In addition, strong team and leadership skills and an active commitment to academic self-governance are essential.

The university will employ professors with clinical duties on a private contract basis (privatrechtliches Dienstverhältnis) according to an agreement on objectives with the State Ministry of Innovation, Science, Research and Technology of North-Rhine-Westphalia. Exceptions are possible in case the appointee holds already a tenured professorship (W2/W3, C3/C4). The university or the university hospital will not offer liquidation proceeds. The clinical duties of the professorship will be settled within a separate contract, a performance-linked salary will be granted.

Applications from female candidates are encouraged. In the case of equal suitability, ability and professional achievement, female candidates will be given preference, provided that the personal qualities of another candidate are not more appropriate for the position. Disabled applicants will be given preferential consideration in the case of equal aptitude. Applications comprising all relevant documents (curriculum vitae, details of scientific career, certificates, list of publications, overview of third-party funds procured, list of courses given, reprints of the five most important publications as well as a research concept) must be sent in electronic form as PDF (max. 15 MB) within 4 weeks of publication of this announcement to Professor Joachim Windolf MD, Dean of the Faculty of Medicine at Heinrich Heine University Düsseldorf, to the following email address: **berufungsverfahren@med.uni-duesseldorf.de**

Please also send the Faculty's "Applicant Questionnaire", evidence of your own teaching activities and a description of your personal teaching portfolio. For further information please see **http://www.medizin.hhu.de/akademische-verfahren/berufungen/aktuelles/aktuelle-ausschreibungen.html** The questionnaire should be submitted electronically together with the other documents.

W230820R

NATURAL
ENVIRONMENT
RESEARCH COUNCIL

Met Office

Head of Earth-system Science and Model Development

£45,486 - £52,706 p.a. pro rata, or in exceptional circumstances £54,283 - £66,725 • three-year fixed-term contract • secondment considered. Location • at least 60% at the Met Office headquarters in Exeter

Take earth-system science to the next level.

We need a high profile scientist to maintain and strengthen the UK's leading international position in Earth-system science. It's a chance to lead our efforts on the international stage and inspire future generations. You'll work with both the Natural Environment Research Council (NERC) and the Met Office on the development and exploitation of a world leading Earth-System Model (ESM).

Part of the Joint Weather and Climate Research Programme (JWCRP), you will lead a core team of scientists and technical experts who will bring together, build, and evaluate an integrated ESM from components developed within the UK. What's more, you'll take a lead role in the development of the science and leadership skills of NERC and Met Office staff and students working on the project.

Imaginative, influential, and innovative, you will bring a sustained track record of high quality research in earth-system science, earth-system modelling, or climate change research supported by a PhD or other doctorate. Skilled in managing major projects dealing with the technical, computational, and diagnostic needs of large modelling systems, you will have proven leadership skills. Your ability to collaborate across subject areas will add new momentum to this field.

Closing date: Monday, 10 December 2012

Interviews to be held on 17 January 2013

For more information and to apply, please visit http://jobs.leeds.ac.uk/ and complete the Vacancy Search Criteria. Job Ref: ENVEE0144

We welcome applications from all sections of the community. All information is available in alternative formats please contact +44 (0)113 343 4146.

UNIVERSITY OF LEEDS

U231242R

WORKING TOWARDS EQUALITY AND DIVERSITY **www.leeds.ac.uk**

University of St Andrews

600
YEARS

600th Anniversary PhD Scholarships

The University of St Andrews was founded in 1413 making it Scotland's first university and the third oldest in the English-speaking world. Over the past six centuries it has established a reputation as one of Europe's leading and most distinctive centres for teaching and research. This is an historic moment for the University as we look back on six centuries of tradition and of belief in the value of people, ideas and excellence. It falls to the present generation to preserve these values for the generations to come. As we stand on the threshold of our seventh century, the University is to invest in the future of the next generation of world-leading researchers by launching a new prestigious PhD scholarship scheme. These anniversary scholarships are available from September 2013 and comprise fees plus an annual stipend of up to £15,000 for three years. These scholarships are available across all the disciplines in the Arts & Humanities, Sciences, and Medicine. Up to 100 anniversary scholarships are available.

Closing date: 18th January 2013.

The scholarships will be awarded on the basis of academic quality and research promise, and are open to all to apply. General advice on how to apply for a PhD place and how to be considered for an anniversary PhD scholarship can be sought from PG Admissions Office by emailing pgrecruitment@st-andrews.ac.uk Specific details on how to apply in a particular discipline may be obtained via the Director of Postgraduates for each School.

The University of St Andrews is a charity registered in Scotland (No SC013532).

University of St Andrews

The University is committed to equality of opportunity

www.st-andrews.ac.uk

U231236R

naturejobs.com

THE LATEST SCIENCE JOBS
ANYTIME, ANYWHERE

Download the free *Naturejobs* app at **nature.com/mobile/naturejobs**

i
naturejobs.com
postdoc
London
Find jobs
Recent searches
postdoc
London
Added 04/07/2012
cell development
Singapore
Added 04/07/2012
assistant professor

Available on the App Store

ANDROID APP ON Google play

naturejobs.com

nature publishing group npg

The University of Edinburgh

The University of Edinburgh is an exciting, vibrant, research led academic community offering opportunities to work with leading international academics whose visions are shaping tomorrow's world.

College of Science and Engineering

Chemistry

Chair/Readership

Salary Scale: Within the Professorial/Readership Scale
Appointments will be made at UE09 (£46,846 - £52,706) and UE10 grade

The University of Edinburgh is seeking to make up to three senior appointments to titular chairs, personal chairs or readerships to be held in the School of Chemistry. Applications are particularly welcomed in the areas of organic chemistry and theoretical/computational chemistry but exceptional candidates in other fields that complement current research activities will also be considered. The School of Chemistry, together with St Andrews School of Chemistry, forms EaStCHEM, the strongest Chemistry Research School in Scotland and amongst the very best in the UK. The School has strength and breadth in fundamental research, strong interactions with industry and world class research support facilities.

The School of Chemistry at the University of Edinburgh holds a Gold Athena SWAN award, in recognition of excellence in science, engineering and technology employment in higher education relating to supporting women in science.

Prospective candidates are welcome to visit the School, and are invited to make contact through the Head of School, Professor Eleanor Campbell FRS (email Chemistry.HOS@ed.ac.uk, telephone 0131 650 4729).

Apply online, view further details or browse more jobs at our website. Ref: 006182N. Closing date 31 December 2012.

Committed to Equality and Diversity

The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336.

www.ed.ac.uk/jobs

U231235R

MANCHESTER
1824
The University of Manchester

Faculty of Medical and Human Science

Institute of Human Development

Lecturer in Metabolomics or Applied Mass Spectrometry

£37,012 - £39,257 p.a. **Ref: M&HS-01931**

We seek a skilled and dedicated expert in Metabolomics or Applied Mass Spectrometry to join the Centre for Advanced Discovery and Experimental Therapeutics (CADET), which has recently been established as a joint venture between The University of Manchester and the Central Manchester University Hospitals NHS Foundation Trust (for further information, see www.cadetmanchester.org/). The central remit of CADET is to use proteomic and metabolomic strategies to discover new therapeutic targets for treating chronic diseases, such as diabetes.

You should have a background in chemistry, biochemistry, pharmacology or a related field with significant experience in small molecule mass spectrometry. You should have a relevant PhD and significant postdoctoral experience or suitable industrial experience. Excellence will be assessed from your academic record or proven industrial achievements. Research will be performed in the context of cross-disciplinary teams of proteomics, bioinformatics and translational medical researchers, and therefore an enthusiasm for collaborating is vital. The University of Manchester has an excellent training programme for new academic staff.

This position is permanent subject to a satisfactory probationary period.

Informal enquiries can be made to: Professor Garth Cooper on 0161 701 0240, email: garth.cooper@manchester.ac.uk or Professor Paul Bishop on 0161 275 5755, e-mail: paul.n.bishop@manchester.ac.uk

For further information and to apply for this vacancy please visit our website.

If you are unable to apply online please request an application form by emailing hrrecruitment@manchester.ac.uk or calling 0161 275 8835 quoting the reference number.

Closing date: 14 December 2012.

The University of Manchester values a diverse workforce and welcomes applications from all sections of the community.

www.manchester.ac.uk/jobs

POSITIVE ABOUT DISABLED PEOPLE

U231240R

"When trying to recruit good scientists for such a remote institutions like ours, in the Canary Is., the high visibility and wide-range exposure that Naturejobs provides has been crucial. I'm very happy to have international candidates who applied to our position from places ranging between St. Petersburg and Cincinnati. Thanks a lot!"

Eduardo Salido, MD, PhD,
Hospital Universitario de Canarias

naturejobs

---

naturejobs

Need help with your career?
Ask the Expert on **naturejobs.com**

- Browse past Q&A
- Get professional advice
- Email your questions

Ask the Expert is located in the **Career Toolkit** section of **naturejobs.com**

naturejobs.com

Grow with us

Over the past twelve years, we've worked hard to craft naturejobs.com into an eminent career resource for scientists. In that time, with your help, we've become the largest dedicated jobs board for the scientific community worldwide.

That's why we're excited to share the new naturejobs.com – with improved functionality that lets you search, save and apply for jobs more quickly and easily.

- ✓ **Easily find jobs with our improved search**
- ✓ **Quickly save job alerts**
- ✓ **Apply faster with our CV upload feature**

Ready? Search over 10,000 vacancies to find exactly the right job for you at:

www.naturejobs.com

Follow us on:

nature publishing group npg

the Beatson
institute for cancer research

Postdoctoral Research Scientist

Starting Salary from £28,500 to £38,000 (Depending on experience)

Initial 3 year fixed term contract (with possibility of up to 2 year extension)

Autophagy and Apoptosis in cancer
Prof Kevin Ryan

The Beatson Institute in Glasgow is one of Europe's leading cancer research centres. It is core funded by Cancer Research UK and supports cutting edge research into the molecular mechanisms of cancer development. The Institute provides an outstanding research environment, underpinned by state-of-the-art core services and advanced technologies.

The Tumour Cell Death Laboratory led by Professor Kevin Ryan is interested in factors that regulate autophagy and apoptosis and how they contribute to tumour development and cancer therapy (Crighton et al. (2006) Cell 126:121-, Bell et al. (2007) JCI 117:1008-, Wilkinson et al. (2009) Genes Dev. 23:1283-, Liu and Ryan (2012) J. Cell Science. 125:2349-). We are seeking a highly-motivated and dedicated postdoctoral researcher to study the role of autophagy in cancer. In particular, you will investigate a pathway connected to autophagy which could be used for the early detection of human cancer. Studies will involve RNAi, confocal microscopy, proteomics and mutational and biochemical analysis of protein function in vitro and in vivo. A relevant PhD is essential and evidence of practical experience in molecular and cellular biology and/or in vivo techniques would be an advantage.

Applications with CV and names of three referees should be sent to Prof Kevin Ryan, The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK or by email to k.ryan@beatson.gla.ac.uk

Closing date for applications: 31 January 2013

CANCER RESEARCH UK

Scottish charity no SC006106

U231237R

---

nature COMMUNICATIONS

Assistant, Associate or Senior Editor (Earth Sciences)

Nature Publishing Group (NPG) is a division of Macmillan Publishers Ltd, dedicated to serving the academic, professional scientific and medical communities. *Nature Communications* is an innovative multidisciplinary science journal, which launched in April 2010. The journal publishes high-quality research from the breadth of the natural sciences, and has already developed a strong track record in these areas. The journal offers rapid publication in an online environment, with Open Access options.

We now have an exciting opportunity for an editor to join our team. The ideal candidate will have a PhD (or equivalent), and preferably postdoctoral experience, in an area of earth sciences. Editorial experience would be beneficial, but is not required. Reporting to the Managing Editor (physical sciences), the successful candidates will play an important role in determining the representation of their field in the journal, and will work closely with the other editors on all aspects of the editorial process, including manuscript selection. This role will also entail significant liaison with the scientific community, the Research Journal Chief Editors (UK and US) and the Editorial Advisory Panel.

This is a demanding and intellectually stimulating position. Broad scientific knowledge and training, excellent literary skills and a keen interest in the practice and communication of science are a prerequisite. The successful candidates must therefore have excellent communication and interpersonal skills. The salary and benefits will be competitive, reflecting the critical importance and responsibilities of this role. This role is offered on a full-time, permanent basis and is located in our London office.

All applicants must be able to demonstrate the right to live and work in the UK in order to be considered for the position in the UK.

Closing Date: Friday 7th December 2012

Applicants should upload a CV, a covering letter explaining their interest in the post and a concise discussion of recent scientific developments that you have found particularly exciting (stating why) to vacancy **NPG/126/12** on http://jobs.macmillan.com/.

nature publishing group npg

IN230944R

Change your environment. Find jobs where you'll make a difference
naturejobs

www.naturejobs.com

POSTDOCTORAL SCIENTIFIC RESEARCH GRANTS FOR THE USA

The English-Speaking Union offers up to three grants a year worth $30,000 each for postdoctoral scientists to carry out research in the USA. The Lindemann Trust Fellowships support research in the fields of **Astronomy, Biophysics, Chemistry, Engineering, Geology, Geophysics, Mathematics and Physics.**

ELIGIBILITY: candidates must be British or Commonwealth citizens, resident in the UK, and have completed a doctorate since September 2010, or be expecting to complete by 30 September, 2013 at the latest. Holders of established research and lecturing posts with a proven record may apply, but the Fellowship should not be regarded as a means of taking up a routine sabbatical. Fellowships may not be held concurrently with any other research grant or scholarship.

LINDEMANN TRUST FELLOWSHIPS 2013-2014

Postdoctoral scientific research grants for the USA

The English-Speaking Union

Registered Charity No: 273136

Closing date: **18 February 2013**
Interview date: **16 April 2013**
For more details contact:
Education Department, The English-Speaking Union
Dartmouth House, 37 Charles Street
London W1J 5ED
T: 020 7529 1550 **education@esu.org**

U230983A

REPROGRAPHICS: CC Media Group, London EC1V 8AR UK and **The Charlesworth Group,** Wakefield, UK.
PRINTED BY: William Gibbons & Sons Ltd, UK; **Publishers Press,** Lebanon Junction, Ky, USA and **Obun Printing Co. Inc,** Tokyo, Japan

TRANSMISSION RECEIVED

A difficult journey.

BY PETER J. ENYEART

Such journeys are permitted only for those who submit to death and rebirth. Eva remembered those words as they put her to sleep, as the bright round light overhead started to spin and then went out.

The company rep had said those words when Eva asked why she couldn't be shipped out to the asteroids in her original body. Oddly poetic for a corporate headhunter. She supposed he thought it sounded more momentous than saying: "It's not worth the time and expense to drag 60 kilograms of meat up out of the gravity well and across 30 vacuous light minutes when we can just radio the data necessary to reconstruct it on the other side."

"By providing your digital signature here," he had continued, "you affirm that you accept the Employment and Transmission Contract and understand its terms. We are required to remind you at this point that once the transmission and reconstitution of your data is complete, the original will be destroyed, in accordance with the law, which allows only one physical copy of a given individual to be in existence at any given time." When she'd hesitated, he'd smiled and put his hand on hers. "It doesn't hurt. You'll be put under anaesthesia before the recording is done. Then you'll just wake up at the other end. I've done it several times myself." She'd resisted the urge to jerk away.

Eva had lost her lab as a result of false charges of academic dishonesty. Subsequent legal wrangling had exhausted her finances but yielded nothing. She was rejected from every remotely technical job she had applied for. Except one. Having nothing left to lose engenders boldness. She signed.

"The company will cover the costs of the basic transmission package, which guarantees a high enough resolution for you to perform your duties, but may entail amnesia, aphasia and partial paralysis, and an increased risk of neoplasia, fibromyalgia, aneurysm and osteoporosis, among other conditions."

"How can you guarantee I'll be able to do the work you need at the other end?"

"The prospective individual will be subjected to physical and psychological evaluations at the work site. Should the prospective individual not be up to spec, or should performance subsequently deteriorate, a replacement will be transmitted at no charge to you from the digital record of your molecular structure stored in our servers.

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

"There is also an upgrade to the high-resolution package available for purchase. Actually, I am pleased to inform you that you have been pre-approved for a company loan that will cover the costs if you lack sufficient funds. The rate is very reasonable, and, conveniently, payments will be automatically deducted from your salary. Would you like to upgrade?"

JACEY

"Bastards!" she whispered, and opened her eyes.

The light above her was still bright but was now an oval. Her head was immobilized, and her wrists and ankles were strapped down. This was not in the contract.

"Congratulations, Eva," said a soft male voice. "You have escaped the cycles of linear death and rebirth." She looked around as best she could but couldn't see anyone. It smelled of bleach.

"What is it with you people and the cult talk?" she responded.

A face came into view. A young man. The bright backlight obscured his features. The voice she had heard before laughed, but didn't come from the man above her.

"Do you know where you are?" the man asked. It sounded like the first voice. Another face came into view, identical to the first.

"I damn well better be on 9 Metis, or I'll be suing for breach of contract."

The laugh again, off to the side.

"A version of you is on 9 Metis …"

A third version of the face now came into view. Realization hit her.

"You pirated me! Who the hell are you?"

"Clever *and* feisty. Yes, I'm glad we chose you."

"Who the hell are you?"

"Calm down." They seemed to take turns talking.

"You tell me where I am, and who you are, now!"

The faces gave each other concerned glances.

"We monitor the System Government's transmissions for individuals who have qualities we admire. When we find one, we make a copy."

Eva was intrigued. "How can you parse that information from the signal without physically reconstructing the whole person? That technology doesn't exist."

"It doesn't exist in the space controlled by the System Government. But the System is decadent. The desire to advance technology is gone, and anything startlingly new is seen as a threat to the status quo. You should know."

She scowled.

"We've been following your work for some time. We were very excited at the prospect of having you on our team."

"So what now?"

"Join us as we develop the technology to destroy the System Government and bring human civilization to the next phase of its development. Or be destroyed as an illegal copy."

"What choice do I have? I'm with you," she said, although her mind was already working on alternative arrangements.

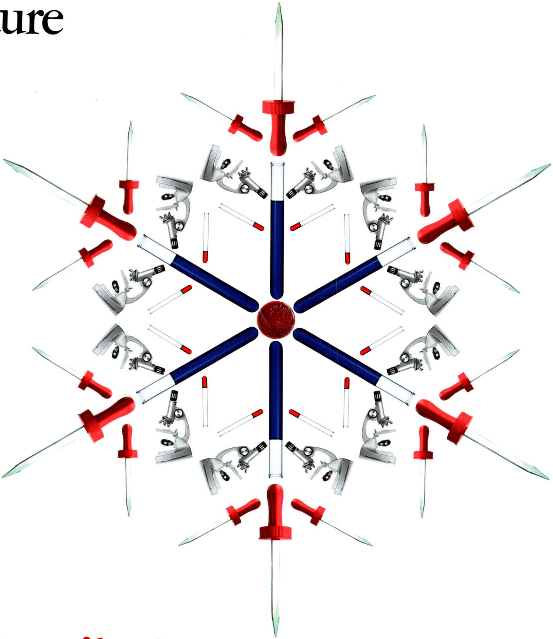
"Excellent. Each of your copies will get a free back-up every six months, and — wait." The faces looked concerned. "Unfortunately, your neurological monitors are showing intense animosity. We'll have to try again with an approach better suited to building your confidence."

"No, wait! You can't —"

They smiled, and a hand patted hers. "It doesn't hurt. You'll be put under anaesthesia and then you'll just wake up again. It happens a few times to everyone until we learn how best to tailor our introduction to the new recruit." She tried to jerk away.

"Bastards!" she whispered, and opened her eyes. ■

Peter J. Enyeart *is a graduate student working towards a PhD in cell and molecular biology at the University of Texas at Austin. He also enjoys sleeping and patent translation.*



nature

Subscribe to *Nature* this Holiday season

50% discount!

Celebrate with *Nature* at our exclusive festive rate and enjoy the following:

- 51 issues of cutting-edge research, global news, perceptive commentary and archive access back to January 1997.
- Unrestricted online access plus multimedia features such as *Nature Podcast*, *Nature Video* and blogs.
- Access via the nature.com app for iPhone®, iPad® and iPod touch®.
- Access to topical supplements including *Nature Insights*, *Nature Outlooks* and *Nature Collections*.

iPhone, iPad, and iPod touch are trademarks of Apple Inc., registered in the U.S. and other countries.

nature.com/holiday12

nature publishing group npg

SIGMA-ALDRICH®

SIGMA Life Science Where bio begins™

More translational and predictive animal models.

Plot your pathway for breakthroughs in neuroscience with knockout rat models from SAGE® Labs.

Biosynaptic.

Advancements in neuroscience are yours to discover with targeted knockout rats from SAGE Labs. Our comprehensive suite of rat models can help move your neuroscience research forward. Map your next breakthrough with smarter rat models for studying neurodegenerative, mood, and affective disorders, including seven new models of autism.

SAGE® LABS
Sigma Advanced
Genetic Engineering

DEPRESSION
ALZHEIMER'S
PARKINSON'S
SCHIZOPHRENIA
AUTISM
PAIN

Explore our collection
of knockout rat models
sageresearchmodels.com

©2012 Sigma-Aldrich Co. LLC. All rights reserved. SIGMA and SIGMA-ALDRICH are trademarks of Sigma-Aldrich Co. LLC, registered in the US and other countries. SAGE is a registered trademark of Sigma-Aldrich Co. LLC. Where bio begins is a trademark of Sigma-Aldrich Co. LLC.